Pattern Recognition Letters xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Probabilistic human interaction understanding
Exploring relationship between human body motion and the environmental context

Kamrad Khoshhal Roudposhti [a,*], Jorge Dias [a,b]

[a] University of Coimbra, Portugal
[b] Khalifa University, United Arab Emirates

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Relationship characteristic
Human interaction understanding
Human movement understanding
Bayesian approach
Hidden Markov Model
Laban movement analysis

ABSTRACT

This paper presents an approach for modeling human interactions based on existent relationship characteristics between body parts motions and environmental parameters. Human interactions properly cannot be identified without knowing the relations between the objects such as human-robot and human–human. During any human interaction, there are many relations between human body parts and others. In this article a general model to analyse human interactions based on the existent relationships is presented. To study human motion properties, Laban Movement Analysis (LMA), a well-known human motion descriptor is used. This work focused on*Relationship*'s component of the LMA to analyse and formulate human activities related to environment. Bayesian approaches are proper classifiers for the mentioned goal, in order to be able to predict, define the existent dependencies, fuse different types of features and also deal with uncertainty. To present the idea, the model was performed to estimate some human movements and activities related to an object like a robot or another person. The result proves the capability of the approach to model and analyse any human activities related to environment using the LMA framework.

1. Introduction

This paper presents a new approach to analyse human activities based on the existent relations between different human body motions properties and environmental parameters. Different kinds of human body motion-based features such as; position, velocity, acceleration, trajectory, etc. have been used to analyse human movements and activities. In addition to those features, the relation between them is very important for analyzing different types of human movements. These relations can be defined in three levels. In the first level, relations are derived from internal body part motion-based paramenters. Then the relations between one person and environment are needed to analyse human activity in a scene. Finally human–human interaction can be estimated by defining the relationship of human motion characteristics of a couple of persons. For instance; in a normal walking action, many features can be obtained, but the harmony between legs and hands is an important feature which leads to action perception.

In another example, suppose one clapping his hands, the "hands touching each other" is the key feature (relation of those body parts). The relation can be defined also between a body and an external object such as reaching to or grasping a cup of water, and also other body such as hand shaking, punching, etc. In this work an approach is presented to analyse all those types of human activities based on the relationships.

This work is inspired from a well-known human motion analysis system (Bartenieff and Lewis, 1980), Laban Movement Analysis (LMA), which was created for choreography and dance notation purpose. The system has a framework that consist of different types of features to interpret human motions. One of those types of features is provides a way to explain the relationship between a human body part motion and other objects (other body parts of itself, environment and other body). Those relationships can be appeared widely, and having a global framework that could be able to explain them as much as possible, is an important challenge that we attempt to obtain it by the LMA system.

Fig. 1 presents a sequence of human movements, which can be analysed as a sequence of human interactions with a robot. Those analysis can be useful in several application, e.g. smart-home, surveillance, human-robot and human–human interaction analyses, etc.

In our previous works, several features based on the LMA concepts were defined to analyse human movements. An approach was presented to use frequency-based features to estimate human

* Corresponding author.
E-mail addresses: kamrad@isr.uc.pt (K. Khoshhal Roudposhti), jorge@isr.uc.pt, jorge.dias@kustar.ac.ae (J. Dias).

0167-8655/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2012.09.021

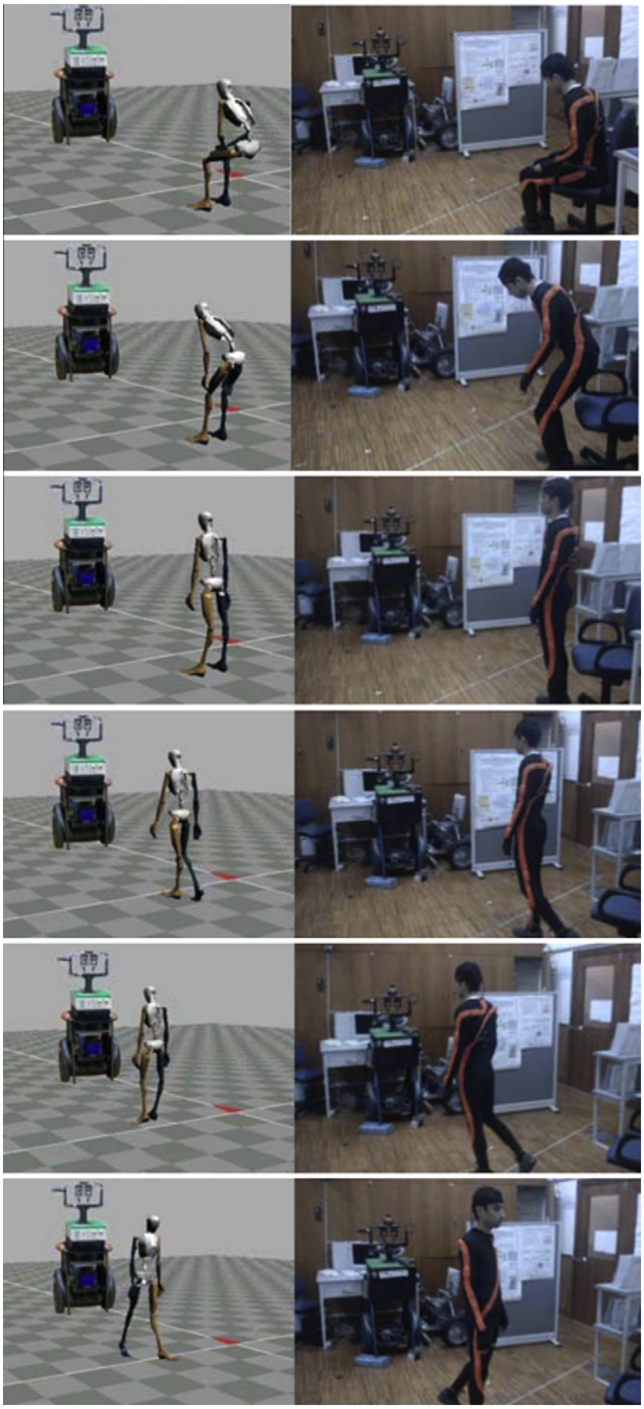

Fig. 1. A sequence of human movements with respect to a robot. The images in the right side are shows a person who wore a motion tracker suit (MVN®). Left side images represent the corresponding skeleton obrained from the motion tracker suit.

movements based on some body parts acceleration signals in (Khoshhal et al., 2010). Then spatial based feature was integrated to improve the previous estimation results in (Khoshhal et al., 2011). The 3D data, which were used on the previous works, were collected by a motion tracker suit. In this work, we attempt to define the relationship's parameters using the previous obtained features, and also environmental parameters. To implement the idea, Bayesian Network (BN) and Hidden Markov Model (HMM) as a Dynamic Bayesian Network (DBN), which are well-known approaches in this area (Poppe, 2010), are used. The capability of generalization of the system, because of using LMA framework, is the most advantageous of this approach.

1.1. Previous related works

Analysing human motion is a prerequisite for understanding any human activities, such as human-robot interaction, human–human interaction, etc. Analysis of human activities can be investigated in different levels. Aaron Bobick in (Bobick, 1997) presented a survey about the different levels of human motion definitions, such as human movement, activity and action. In Bobick's terminology, the movements are the lowest level of human motions which do not need any contextual and previous knowledge to be identified. To understand human action or behaviour we need to have a sequence of human movements or states with respect to the environment or scenario.

There are many different kinds of human movements that researchers try to recognize. Each of them can be identified by position variations of one or more involved human body parts, without the need to know the underlying movement concept. Thus, researchers whose research concerns to estimate human dynamic characteristics, try to track body parts in a sequence of captured data such as image sequences.

An accurate human movement tracking is essential for a precise human interaction analysis. Feet, hands, head and face are the most important body parts to track in related applications. Some examples can be given such as; gait recognition, where the dynamics of leg motions are important, facial expression recognition, the focus lies solely on the face, while in gesture recognition, all parts of body can be involved, where hands and head stand out as the most important ones.

There are several surveys about human motion analysis such as Aggarwal et al. (1994), by Aggarwal et al., which covered various methods used in articulated and elastic non-rigid motion. Cedars and Shah in (Cedras and Shah, 1995) presented an overview of methods for motion extraction, in which human motion analysis was illustrated as action recognition, recognition of body parts and body configuration estimation. Gavrila in (Gavrila, 1999) described a work in human motion analysis in terms of 2D and 3D approaches. Pentland in (Pentland and Liu, 1995) touched on several interesting topics in human motion analysis and its applications. Moeslund and Granum (Moeslund and Granum, 2001; Moeslund et al., 2006) published a couple of surveys of computer-vision-based human motion capture problems (initialization, tracking, pose estimation and recognition). Wang et al. in (Wang et al., 2003) provided a comprehensive survey of research on computer-vision-based human motion analysis (human detection, tracking and activity understanding). Poppe in (Poppe, 2007) described the characteristics of human motion analysis, and the study divided the analysis into modeling and estimation phases.

As Bobick said, to analyse human activities, we need to know the underlying movement concept (Bobick, 1997). It means to understand human activities, we need to find the relation of between human movements and environmental parameters. Since there are infinite relations that can be appeared in human activities, thus, researchers always define some specific relationship properties to present their methods. For instance; Rao et al. in (Rao et al., 2002) presents a computational representation of human action to capture the changes using spatio-temporal curvature of 2-D trajectory of hands. Then in the experiment part, some activities such as; pick up an object from the floor, and put it down on the desk, were defined. Thus, It can be realized that we need a general framework that to be able to analyse those kind of relation parameters in mid-level. LMA is a system that can be used as the mid-level features of human motions. To have this mid-level, sev-

Please cite this article in press as: Khoshhal Roudposhti, K., Dias, J. Probabilistic human interaction understanding. Pattern Recognition Lett. (2012), http://dx.doi.org/10.1016/j.patrec.2012.09.021

eral works have been done for several types of features, based on the LMA systems (Foroud and Whishaw, 2006; Khoshhal et al., 2011; Khoshhal et al., 2010; Rett et al., 2008a; Zhao and Badler, 2005). All of them were not explored in relationship component parameters of LMA, but in *Effort*, *Shape* and *Space*.

For implementing the idea, Bayesian approach has been used. A Bayesian approach presents many advantages on using prior knowledge and modeling the dynamic dependencies between parameters of object states. In the related fields, this approach is popular and researchers have been keen on applying it. For example; in (Rett et al., 2008b) Rett applied a general BN framework for analyzing human motions based on the LMA concept for a human machine interaction application, however in that work the relationship parameters were not investigated. Ryoo and Aggarwal (Ryoo and Aggarwal, 2006; Ryoo and Aggarwal, 2008) presented an approach for human action modeling by using a number of BNs to recognize the poses of body parts and a DBN to analyse human activities based on using 2D data.

There are several reasons to use the Bayesian approach in the mentioned applications: Bayes theorem is valid in all common interpretations of probability, can represent and solve decision problems under uncertainty, is a common approach to predict, an explicit approach to follow states, does not need large amounts of data to be trained, and it is able to fuse different types of data in one model.

As it was mentioned, a BN can model the dynamic dependencies between parameters. These dependencies between parameters somehow play some of the relationship roles, but it depends too much on the learning process. The point is, the relation parameters were not defined explicitly in the model, thus it is very sensitive to the learning process. Thus by considering those relationships properties such as the input features of our model, more reliable results can be achieved. Finally by connecting human body motions to the scene, which is one of the *Relationship* component purposes, the probabilistic model for analysing human interactions are explored.

2. Features extraction

Human motions consist of a number of features which can be defined and extracted in different domains. In the previous works, frequency-based features (Khoshhal et al., 2010) and spatial-based feature (Khoshhal et al., 2011) were extracted to analyse human movements. The results showed that by conjugating of the both types of features, the classifier's efficiency significantly was improved. Thus, we used those features relying on existent relations between human body parts motions and environmental parameters. Based on the relationship definitions which are described in the next section, infinite relations can be defined. Therefore to perform an experiment and depend on the activities, it is needed to collect the best features which can disclose differences of the activities perfectly.

2.1. Body motion based features

In this step, we estimate some general human activities, like *walking*, *running*, *siting*, *rising*, *falling down* and *standing,* which can be extended to more types of movements, without using environmental parameters. Thus, we attempt to define the relationship parameters between body parts of a person.

Relationship characteristics are very wide, and play an important role to performing any activities. For instance; in walking types movements, usually there is a harmonic motion in hands and feet related to the body center. Those harmonic motions can be estimated by frequency-based properties (Cheng et al., 2004). The frequency-based features can be extracted by Fast Fourier Transform (FFT) and Power Spectrum (PS) techniques which are the known approaches (Ragheb et al., 2008). By exploring among of different collected signals (trajectory, velocity, acceleration) from body motions, the acceleration signals of human body parts related to the body center were selected (based on the previous work (Khoshhal et al., 2010)), and then FFT and PS signals of the acceleration are extracted (Khoshhal et al., 2010) (see Fig. 2(top-left) and Fig. 2(top-right)).

As (Ragheb et al., 2008) mentioned, power of the PS signals for human motions usually is high in low frequency domains. Thus, based on the previous work (Khoshhal et al., 2010), the peaks of PS signals in the first four frequency sub-domains (1–10 HZ, 11–20 HZ, 21–30 HZ, 31–40 HZ) for different movements are collected as the low level features (LLFs). Other frequency domains data can be achieved, but in these kinds of applications those selected domains are more representative and sufficient. Fig. 2(left-down) presents a histogram of the frequency data for a body part in different human movements. This kind of histogram for each of the selected body parts is generated.

The frequency-based data, which are obtained from acceleration signal, are the proper features to distinguish between similar movements like walking and running, or sitting and falling down, but not for the others, which have difference in terms of spatial property, like sitting and rising, running and falling down. Thus some other useful features to distinguish those movements are the relation of two objects like head and feet in terms ofthe relation between those body parts during the different movements. For instance; in sitting and falling down movements, distance of those body parts reduces and in a rising movement there is an opposite situation. However in standing or walking, there is no considerable changing in this aspect. Thus the difference of distance between head and feet during a movement can be obtained:

$$\Delta D = \sum_{i=2}^{n} \left(\left(X^i_{obj1} + X^i_{obj2} \right) - \left(X^{i-1}_{obj1} + X^{i-1}_{obj2} \right) \right), \tag{1}$$

where X^t_{obj1} and X^t_{obj2} deNote 3D position of the two objects *obj*1 and *obj*2 at frame *i*, respectively. *n* denotes the number of frames inside of a window signal, and ΔD denotes the difference distance between the objects during the window signal in the meter unit. The point is, when one of the objects passed through another during of the window frame, the equation will calculate the difference distance if it is more in reaching or spreading state.

2.2. Environmental based features

To implement this part, couples of scenarios are proposed to present more visible the idea;

- First scenario, which includes a static robot agent that people can interact with that, is defined. Some relevant activities were performed like, *reaching*, *spreading*, and *passing*.
- Second scenario is about a couple of person interaction. In this scenario, there are not only previous activities, but also more complex ones (such as *following*, *handshaking* and *pushing*) which need more input features to be estimated.

The features in this level are differences of distance between two objects during human activities. For the first scenario the difference of distance between a person and a robot using the Eq. 1 is necessary. In the second one we need also the distance between the hands of the couple of persons.

3. Relationship modeling

Laban Movement Analysis (LMA) is a well-known method for observing, describing, notating, and interpreting human movement (Bartenieff and Lewis, 1980; Badler et al., 1993). The theory of LMA consists of several major components, though the available literature is not in unison about their total number. The works of Norman Badlers group (Badler et al., 1993; Zhao and Badler, 2005) mention five major components; *Body*, *Effort*, *Space*, *Shape* and *Relationship*.

Each of those components describes human motion in different aspects.*Body* describes human body part position situation, *Space* interprets direction of human body parts motions, *Effort*or dynamic explains how a body part motion is performing (for instance; if it is sudden or sustained),*Shape*has another aspect of the human motion's interpretation, and explains human body shape (as a blob) deformation during a movement. Several works were published around these components, but there is not any attempt to analyse human activities in relationship's component aspect which could explain human body parts motion relations to other parts of the body, another body or environment parameters (Rett et al., 2008a).

The *Relationship* component of LMA has several parameters that categorized different types of possible relations. As mentioned by Hutchinson (Hutchinson, 1974) these parameters are named such as; Addressing, Nearness, Contact or Touch, Support or Carry, Enclosing or Surrounding, Toward and Away, and Facing. Each of those parameters can be in three situations; Passing, Retention and canceling of the relationship.

In the mentioned application, to analyse a person's simple movements and activities (interact with a static object (see Fig. 1), such as; reaching, passing and spreading, and a couple of persons activities such as; handshaking, following and pushing), some of those parameters which are more representative are used to present the approach. The approach allows us to generalize the system, however for different activities, some of these Relationship's parameters are more sufficient. The performed relationship parameters are described as follows (Hutchinson, 1974):

- *Toward and Away:* A performer may gesture toward or away from a part of his body, another person, an object, or a part of the place. For instance; left hand of a person moves toward other left shoulder, head moves to down in sitting movement, and approaches to or moves away from his/her partner.
- *Passing, transient relationship:* Each of the relationships, addressing, nearness, touching, etc., may occur in passing, this is, the relationship may be established, momentarily sustained, and then relinquished. For instance; right hand passes near the left hand, one person passes near a robot.
- *Retention of a relationship:* When a relationship retains for more than the moment usually depends on what comes next, and when no obvious cancellation occurs, it is expected to remain. For instance; keep the hands near each other, the box is to be kept in the hands, keep the object near to the person.
- *Contact, Touch:* When a part of body is active in producing a touch or contact to another part, an object, or another person. For Instance; Hands touching opposite elbows, Hand-shaking of a couple of persons.

3.1. Relationship's component modelling

Several properties for human activities can be defined in *Relationship*'s concept. As it was mentioned before and to simplify the system, some of those properties are modeled for the performed activities, as following subsections.

3.1.1. Passing and Retention relationship

These properties can be used between every two objects. In this model, the objects are the body parts. During each human activity, there are different motion signals between body parts. For instance, in a walking type movement, there is the same motion signal between opposite sides of hands and feet in the same moment. To analyse those signals, characteristics of passing and retention relationship are used between each body part related to a reference point like body center (if a body part is in the *passing* or *retention* relationship with the body center). If the frequency-based quantities, which are extracted for each of the body parts movements, are more than a specific threshold, it means that there is a passing status related to the body center, thus it states *passing* relationship, otherwise states*retention* ones. Based on the training dataset, some thresholds are defined to discretize the frequency-based quantities (Khoshhal et al., 2010):

$$f^i_{pb} = \begin{cases} No & Max\left\{f^i_{pb}\right\} \leqq 20 \\ Low & 20 < Max\left\{f^i_{pb}\right\} \leqq 150 \\ Medium & 150 < Max\left\{f^i_{pb}\right\} \leqq 1000 \\ High & Max\left\{f^i_{pb}\right\} > 1000 \end{cases}.$$

where $Max\left\{f^i_{pb}\right\}$ denotes a frequency-based coefficient for a body part (pb) (Head, hands and feet), in ith frequency sub-domain where

$$i = \begin{cases} 1 & 1 \leqq f \leqq 10\ \text{Hz} \\ 2 & 10 < f \leqq 20\ \text{Hz} \\ 3 & 20 < f \leqq 30\ \text{Hz} \\ 4 & 30 < f \leqq 40\ \text{Hz}sHZ \end{cases}.$$

Thus for each body part, one variable with a couple of states is defined as following:

$$Pass_{pb} \in \{passing, retention\} \tag{2}$$

Fig. 2(right-down) presents an example of the output of the $Pass_{pb}$model for all body parts during a sequence of movements. That diagram shows the *passing* state probabilities of a human body parts during a specific movement.

3.1.2. Toward and Away relationship

They can be defined in two different spaces; between body parts, and between a person and a robot or other person. By having Eq. 1, *Toward* qualities between any two objects, can be estimated by discretizing of ΔD using a couple of thresholds, which are obtained by observing among the several experimental data set:

$$Toward = \begin{cases} N & \Delta D > 0.3\,m \\ S & 0.3 > \Delta D >= -0.3\,m \\ P & \Delta D < -0.3\,m \end{cases}, \tag{3}$$

where P, S and N denote *Positive*, *Still* and *Negative* qualities of *Toward*'s property, respectively. These properties are used between head and feet as two body parts of one person (ΔD_b), and a person and a robot, or between a couple of persons as two objects in a scene (ΔD_e). Thus there are two types of *Toward*'s variables with the three states. We applied different thresholds on the model (0.1 to 1.0 m), and the mentioned one provided the best result based on our collected data.

3.1.3. Contact or Touch relationship

Some of activities will appears by contacting two objects, for instance; handshaking, grasping a glass, pushing, kicking a ball, etc. For modeling these variables, *handshaking* and *pushing* actions were selected. *Contact* qualities can be estimated by using Eq. 1 for the two interested objects (e.g. two hands for handshaking), and discretizing of the equation output, using one threshold, which are obtained by observing among several experimental data sets:

$$Contact_{pb1-pb2} = \begin{cases} Connected & \Delta D_c \leqq 0.1\,m \\ Disconnected & \Delta D_c > 0.1\,m \end{cases} \tag{4}$$

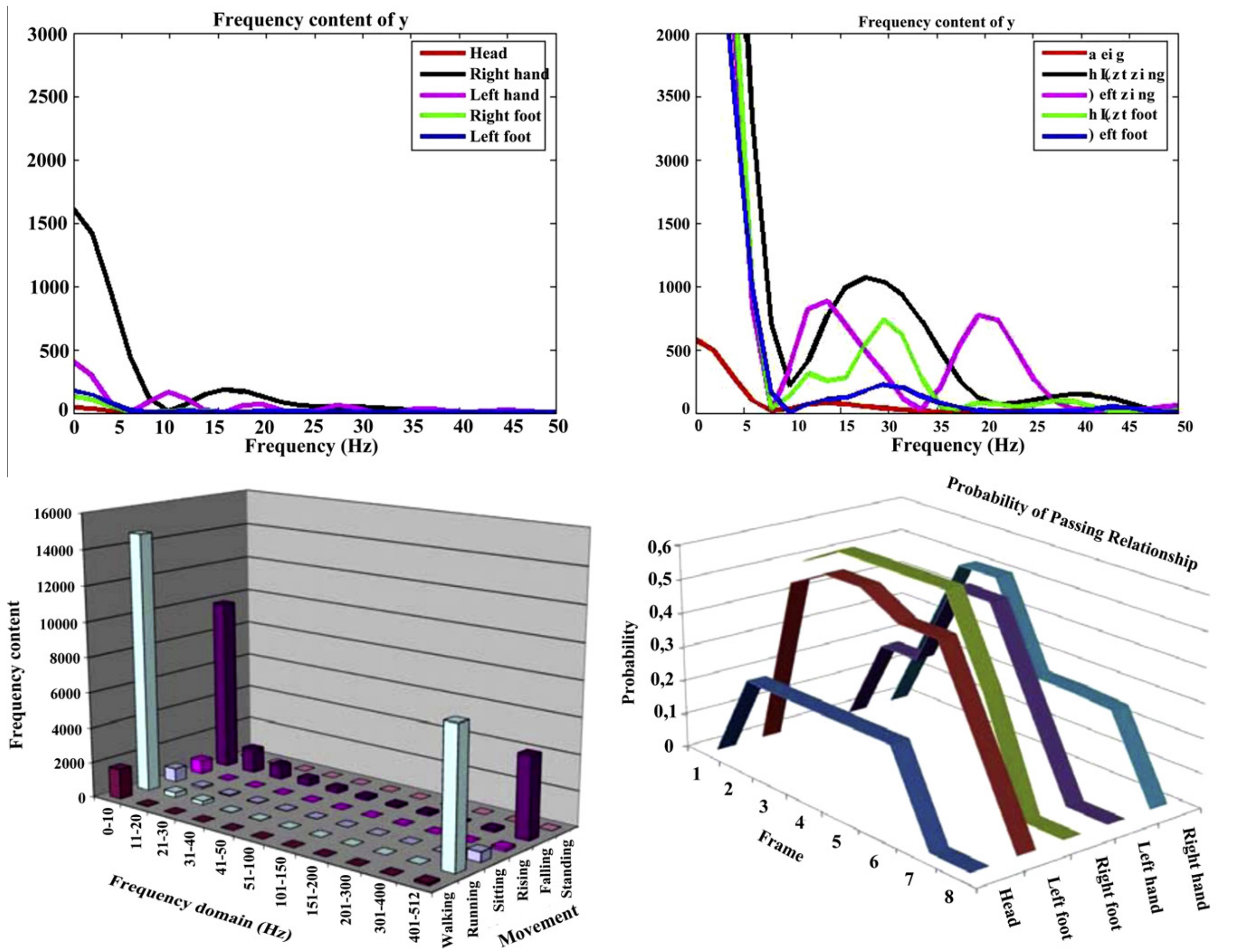

Fig. 2. (Left-top) and (right-top) diagram present PS results of some body parts acceleration signals for walking and running movements respectively. (left-down) presents an histogram to show the frequency-based features of one body part for different movements in different frequency sub-domains. (right-down) *passing* probabilities diagram for all body parts during a sequence of frames.

where $pb1$ and $pb2$ denote hands of the two persons, and ΔD_c denotes the difference distance in contact space. There are two possible states; *Connected* and *Disconnected*, discretized by a threshold 0.1 meter difference distance between two objects.

3.2. Human movement and activity model

By combining the *Toward* variable, defined between two body parts and *pass* variable, the human movements are estimated (see Set.5).

$$Movement \in \{walking, running, sitting, rising, falling, standing\} \quad (5)$$

Finally by using the other variables between person and robot or other person and the *Movement* variable, the performed human activities are estimated. The activity states are:

$$Activity \in \{reaching, spreading, passing, handshaking, following, pushing, other\} \quad (6)$$

Toward's property between human and robot or other person, is the proper feature to analyse *reaching* and *spreading* activities. The *passing* activity can happen when the *Toward* variable states *Still*'s state while a person is walking, but when both persons are walking then *following* activity is happening. For *handshaking* action both persons are in *standing*'s state and the *Contact* variable is in *connected* state. In *pushing* action, however there is a *connected* state also, but one of the persons will perform fall down movement at the end. There are more possibilities to define more activities by having those movements and the environmental parameter states. For modeling the idea, it was not supposed to implement a complex model, but to present an approach which can be easily generalized. Fig. 3(a) presents the idea in a scene. There are three objects (two persons and one robot) in the scene. As can be seen in the figure, depend on the situation of each of them and between each couple of objects, one specific activity can be estimated.

4. Modeling based on Bayesian framework

Bayesian approach is a popular and well known method to classify human motions and activities (Ryoo and Aggarwal, 2008; Rett et al., 2008a). A Bayesian Network (BN) is a suitable method for dealing with variable dependencies and uncertain data, learning with a small bunch of data, and fusing different types of features. HMM (as a DBN) is another popular approach, which is used also for those kind of applications. HMM works when the defined scenarios are based on a sequence of states. In this paper, both approaches are implemented for analysing human activities.

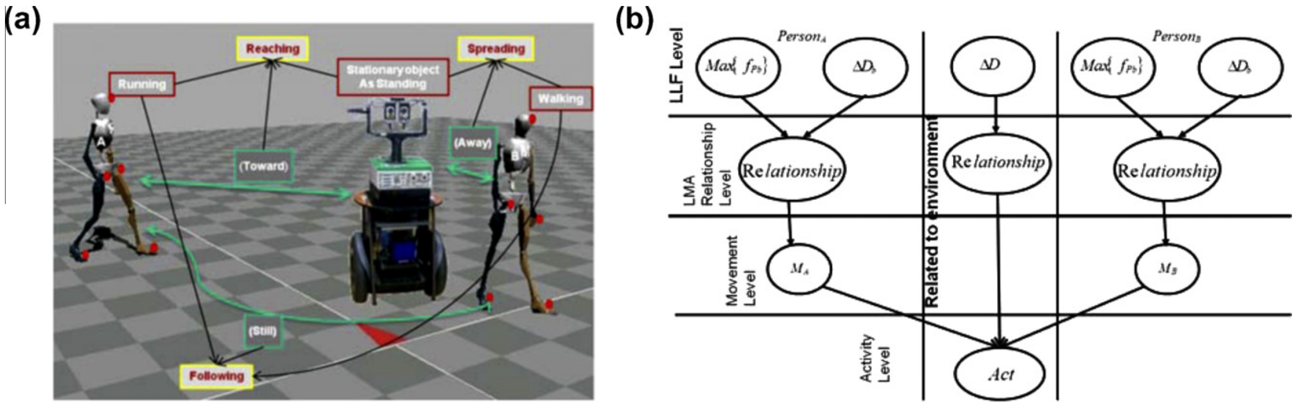

Fig. 3. (a) A scenario of the system. (b) The global Bayesian model for all the process of analysis in four layers, and two different spaces (related or not related to the environmental parameters).

4.1. Bayesian network

In the model (see Fig. 3(b)), as it was mentioned before, there are both frequency and spatial based features in parallel for different properties. Frequency-based features for each body part (pb) are used for the *Passing/retention* relationship of the same body part:

$$P\left(Pass_{pb}|\prod_{i=1:4}Max\left\{f_{pb}^{i}\right\}\right)=\frac{P(Pass_{pb})\prod_{i=1:4}P\left(Max\left\{f_{pb}^{i}\right\}|Pass_{pb}\right)}{\prod_{i=1:4}P\left(Max\left\{f_{pb}^{i}\right\}\right)} \tag{7}$$

The probability of *Toward/Away* relationship between body parts can be obtained by:

$$P(Toward_b|\Delta D_b)=\frac{P(Toward_b)P(\Delta D_b|Toward_b)}{P(\Delta D_b)} \tag{8}$$

and the probability of similar property but between a person and an external object (such as a robot or another person) can be achieved by:

$$P(Toward_e|\Delta D_e)=\frac{P(Toward_e)P(\Delta D_e|Toward_e)}{P(\Delta D_e)} \tag{9}$$

and for the *Contact* property:

$$P(Contact|\Delta D_c)=\frac{P(Contact)P(\Delta D_c|Contact)}{P(\Delta D_c)} \tag{10}$$

There are two other levels of analysis, that one of them is movement's level which is not related to the environment parameters. The other one is activity level which can be analysed by finding the connection between human movement and the scene information. The free-context based movement model is defined as:

$$P\left(M|Toward_b\prod_{pb=1:n}Pass_{pb}\right)=\frac{P(M)P(Toward_b|M)\prod_{pb=1:n}P(Pass_{pb}|M)}{P(Toward_b)\prod_{pb=1:n}P(Pass_{pb})} \tag{11}$$

In the activity level, there are not only each human movement probability but also the environmental parameters:

$$P(Act|Toward_e\,Contact\,M_AM_B)=\frac{P(Act)P(Toward_e|Act)P(Contact|Act)P(M_A|Act)P(M_B|Act)}{P(Toward_e)P(Contact)P(M_A)P(M_B)} \tag{12}$$

where M_A, M_B and Act denote movement state of person A and B, and activity, respectively, and n denotes the number of body parts.

Fig. 3(b) presents the dependencies of those all different variables (LLFs, *Relationship*'s component, movement, activity) in two different space (related and not related to the environmental parameters) in one model.

Fig. 4 presents two examples of the model results in different steps. First scenario is about pushing activity. Fig. 4(a) and and Fig. 4(b) present the both persons movements classifier results, and the Fig. 4(c) shows the trajectories of them in the scene. Fig. 4(d) presents the results of the activity classifier. Another scenario shows handshaking activity. Fig. 4(e) presents a sequence of three images of the related activity. Fig. 4(f) shows the model of the same scenario that obtained by the motion tracker suit, and Fig. 4(g) Fig. 4(h) Fig. 4(i) and Fig. 4(j) present results movement classifier of person A and person B and the trajectories of both persons in the scene, and finally activity results, respectively.

4.2. Hidden Markov Model

HMM as a DBN is a common approach to estimate human activities based on a sequence states. In this work concurrent HMM is implemented which is described in more detail in (Khoshhal et al., 2011). A concurrent HMM is composed of several HMMs (see Fig. 5(a)), and each one describing one class. The inputs of the model are the all probabilities of the both persons movement classes, and the probabilities of the relationship parameters, obtained by the performed BN in the *Movement* and *Relationship* layers. In every sequence of data, the activity class, which has most probability of its corresponding HMM, is the output of the model (see Fig. 5(b)). This is performed by finding the HMM *Act*: that maximizes $P(Act|O_{t-n}\dots O_t)$ where the O_t denotes observation data at time t.

5. Experiments

To obtain the input data (3D position and acceleration of body parts such as hands, feet, head, etc.) a body motion tracker suit (MVN®suit) (Khoshhal et al., 2011) was used. The suit consists of several inertial sensors which attached in several human body parts. There is an interface (which belongs to the suit) to collect and analyse those sensory data and estimate human skeleton. The suit uses a global reference in the scene as a start point. Thus the distance between human body and others such as another person or a robot, during the activities can be calculated (for more information see http://www.xsens.com).

Several different human movements and activities were performed. One second's window which shifts half second, during each record is defined to feature extraction's process of the all

Please cite this article in press as: Khoshhal Roudposhti, K., Dias, J. Probabilistic human interaction understanding. Pattern Recognition Lett. (2012), http://dx.doi.org/10.1016/j.patrec.2012.09.021

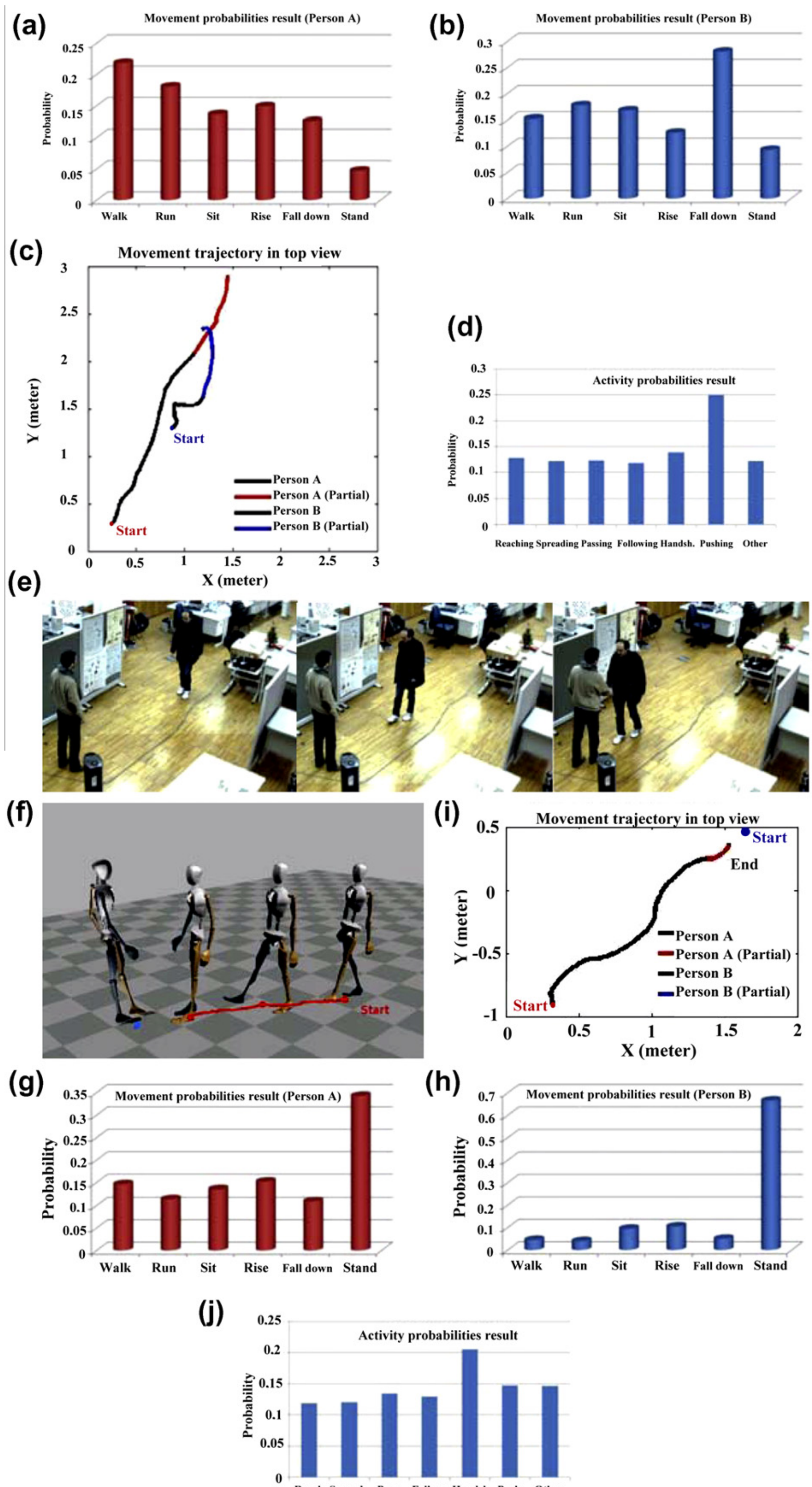

Fig. 4. These diagrams present two samples; first is a *pushing* activity scenario, the persons trajectories were shown in (c), and movements classification results in (a and b) and activity in (d). Second scenario presents handshaking activity. (e) shows a sequence of three images to present a normal handshaking activity. (f) presents an example handshaking scenario which obtained by the motion tracker suit. (g) and (h) shows the movement classifier results of both persons. (i) shows the trajectories and (j) shows the related activity classifier results.

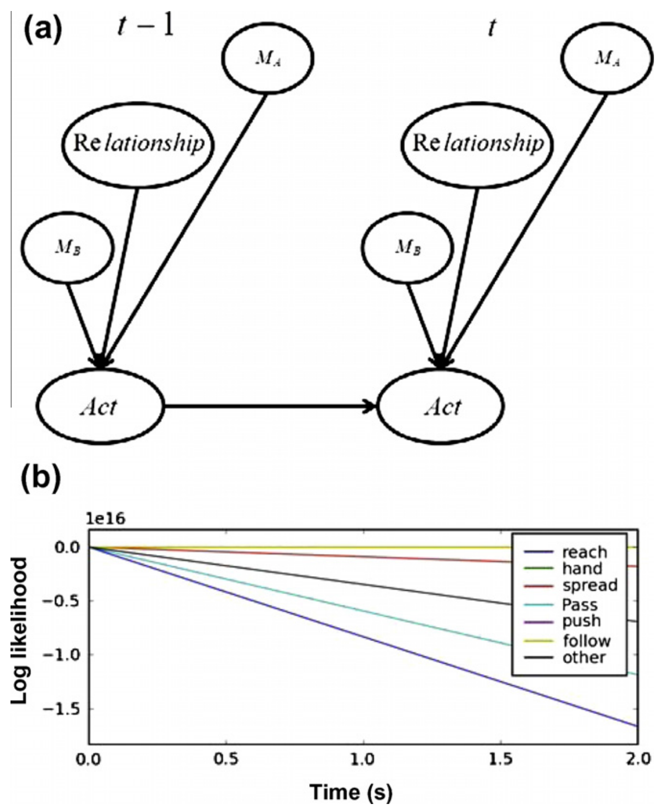

Fig. 5. (a) Structure of the HMM to classify human activity (*Act*) by having several inputs; movements type of person A (M_A) and person B (M_B), and Relationship parameters related to environment (*Relationship*) as the *Toward$_e$*and the *Contact* components, at time t and $t-1$. For each class of *Act* we have one HMM of that. (b) A diagram shows the log-likelihood of each class during the sequence of data. The log-likelihood data are the output of the model. Thus the class which has more log-likelihood than the others will be the activity model estimation result.

performed movements. Two series of data are achieved; 3D human body parts positions related to the body center, and the person's 3D position related to a global reference which the person supposed to interact with, in the scene. A free Bayesian toolbox which provided by Kervin Murphy and Berkeley (Murphy et al., 2005) is used to implement the BN model.

The BN estimates the high probability of the person movement and activity states in each second. Table 1 presents the classification results in movement level, as inputs of activity level of the model. Those data are obtained from around 100 sequence of human movements with different durations (first half of the data is used in learning and others in classification process) and then vice versa (the first half of data is used for the classification process and the others for the learning).

To be able to have a comparison with the stat-of-the-art, we attempt to find the related works in 3D based analysis which is the main characteristic of these works, however the experimental set-up are not the same, and it's not quite fair comparison. In (Holte et al., 2011), 3D data-based classification results for different

Table 1
Human movement classification result.

	Walk	Run	Sit	Rise	Fall	Stand	%
Walk	63					2	96.92
Run	1	72			2		96.00
Sit			46		2		95.83
Rise				34		1	97.14
Fall		1		1	26		92.85
Stand						155	100

number of human movements can be seen. The overall result for classification of six human movements was maximum 89.58%. In this aspect our model provided better performance with 96.45% accuracy. in (Turaga et al., 2008) very accurate classification results (overall 98.78%) were estimated, but their performed movements are quite different in spatial aspect. For instance, there is no running movement in their movement classes which can be easily confused with walking movement for this kind of models. Probably that is the reason that their walking movement's class always estimated 100%. As can be seen in (Holte et al., 2011), their less accurate results were around between walking and running movements.

To present the multi-layer classifier results more visible, Figs. 6 and 7 show probability diagrams of the both classifiers of movement and activity in a sequence of steps. In the first sample of Fig. 6 (first column), one person is in walking and another in running movement state. These two persons will reach in the end and do handshaking activity, but in the 6th frame we see the pushing result activity, because of the person who was in running state in previous step, that usually happens more in pushing than handshaking activity in a normal scenario. In the second column, one person runs and pushes other. In Fig. 7 and in the first column presents a scenario which shows a person how walks and passes near of other person how is in standing state. Finally the second column shows two persons in running state, but one of them is faster than another, thus in the activity results show the person is going away from other one.

Table 2 presents the result of the model for human activities level using the BN model. The result shows that our model for handshaking activity is not accurate as others. The reason is that the related features is not sufficient enough to distinguish between handshaking and pushing activities.

For implementing the HMM approach, an interface described in (Khoshhal et al., 2011), is used. The input data in this case are several sequences of the observations which consist of both of persons movement states and relationship parameters probabilities. For each class, several sequences data for leaning and for classification process, are collected.

Table 3 presents the result of the model for human activity level using the HMM in the last layer. The result shows that the model is more reliable than previous one, with less false detections. It can be improved by using other relationship parameters (Hutchinson, 1974). The advantage of the HMM approach is using a sequence of observations, which is relevant to the activity definition (Bobick, 1997). The results also proved the Bobick's terminology.

A general discussion and comparison between the existent methods for human activities understanding was presented in (Poppe, 2010). In this work, we presented two most popular methods (BN and HMM) in the related area, to show their results in the mentioned framework also. Each of them has own advantageous, for instance; however HMM needs previous state knowledge, but it gives better results than the BN approach. We believe that HMM shows better performance for complex activities.

By having the obtained information by the models, it is possible to analyse more complex human activities, like rubbery (when the rubber performs first reaching and then spreading activity in running movement state's), fighting (reaching in running or falling down movement's states), etc. It means, these relationship's parameters can assist us to analyse even more complex human activities in different concepts and applications.

6. Conclusion and future works

This paper proposed a new approach which allows us to analyse any kind of human activities through the relationship concept.

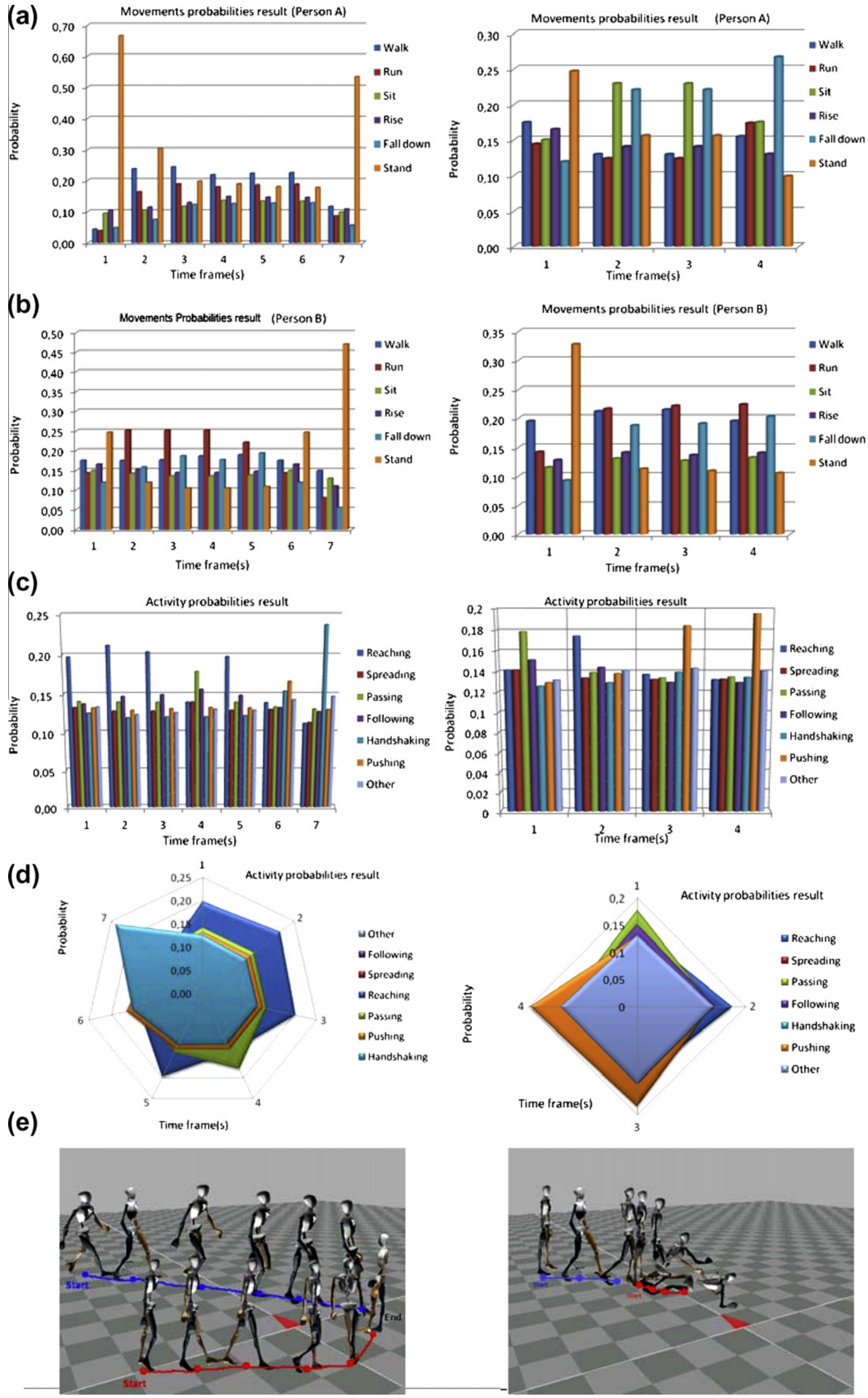

Fig. 6. Two sequences steps of changing the human activity states from two human movements in a scene. (a) and (b) present the both human movements classifier results, (c) and (d) show the activity classifier results in different schema, and (e) presents scenarios (Handshaking and Pushing) in different steps.

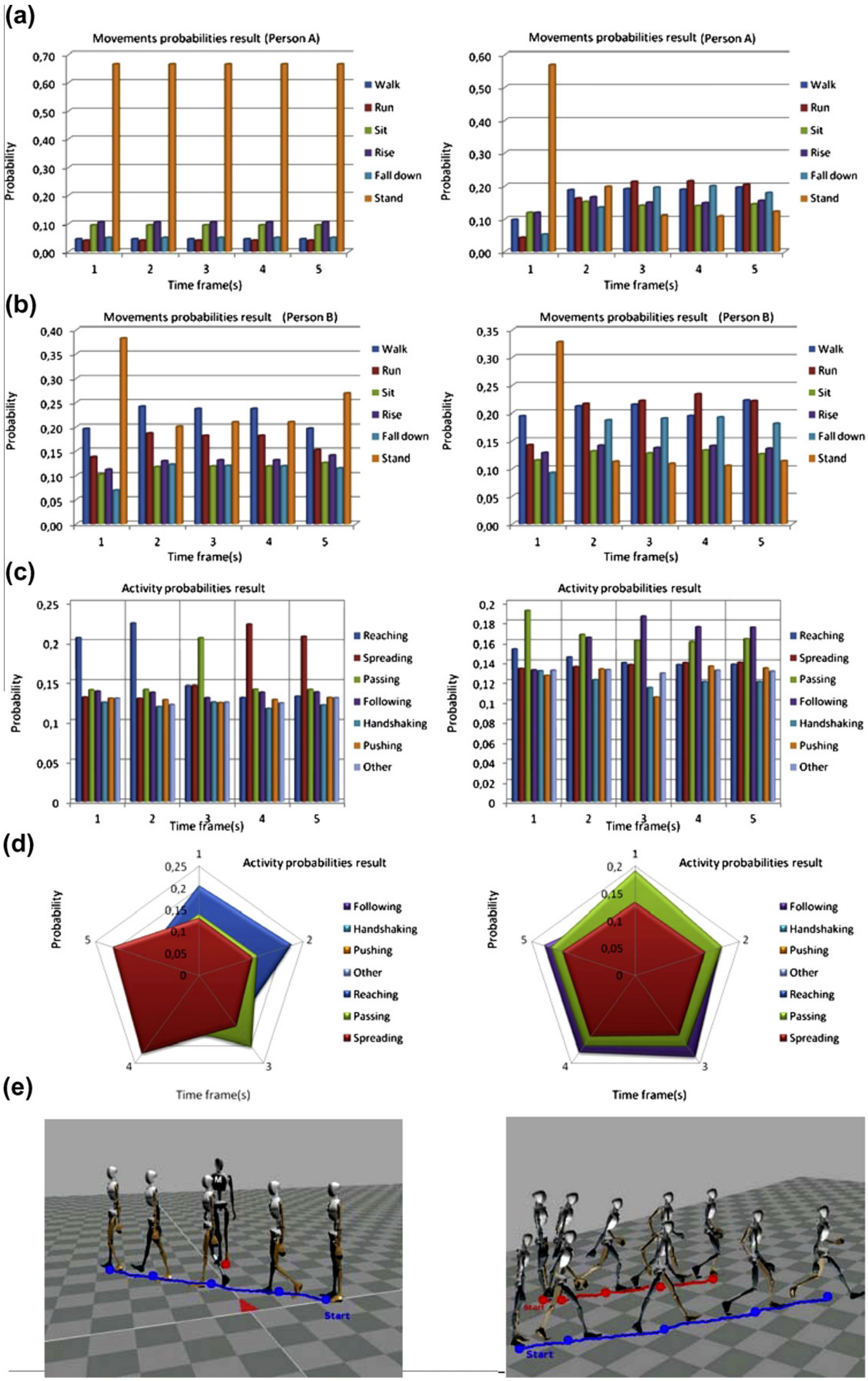

Fig. 7. Two sequences steps of changing the human activity states from two human movements in a scene. (a) and (b) present the both human movements classifier results, (c) and (d) show the activity classifier results in different schema, and (e) presents scenarios (Passing and following) in different steps.

Table 2
Human activity classification result using the BN which can be seen in Fig. 3(b).

	Rch	Spd	Pas	Flw	Hsk	Psh	Oth	%
Rch	133	0	2	0	0	0	7	93.66
Spd	0	141	6	0	0	0	5	92.76
Pas	8	1	127	9	0	0	3	85.81
Flw	1	1	6	50	0	0	3	81.96
Hsk	3	0	0	0	53	11	1	77.94
Psh	0	2	0	0	3	68	3	89.47
Oth	1	1	4	1	1	2	211	95.48

Table 3
Human activity classification result using the HMM as can be seen in Fig. 5.

	Rch	Spd	Pas	Flw	Hsk	Psh	Oth	%
Rch	45	0	1	0	1	1	0	93.75
Spd	0	45	1	0	0	2	0	93.75
Pas	3	0	43	0	1	1	0	89.58
Flw	0	0	0	45	0	1	0	97.83
Hsk	0	1	0	0	42	5	0	87.50
Psh	0	0	1	1	2	44	0	91.67
Oth	0	0	0	0	1	1	55	94.49

There are infinite relations between body part motions and environment in daily human activities. Using *Relationship*'s component of LMA is the key of the system to analyse any human activities specially in terms of interaction with the environment. In this work, we attempted to computerize the component (*Relationship*) for some performed activities.

A Bayesian network was defined to develop a model which can deal with uncertainty and fuse different types of features to classify the movements. In activity process layer, BN and HMM approaches are applied. Results of both classifications show that HMM approach can be more reliable for analysing human activities based on a sequence of states. In the future, we intend to investigate *Relationship* component for interpersonal activities in social aspect to improve our recent work (Khoshhal Roudposhti et al., 2012).

Acknowledgment

This work has been supported by Institute of Systems and Robotics from University of Coimbra, Portugal, and Khalifa University, Abu Dhabi, UAE. Kamrad Khoshhal Roudposhti is supported by the Portuguese Foundation for Science and Technology (FCT)(SFRH/BD/70640/2010). The authors thank Seyed Jafar Hosseini, Luis Santos and Hadi Aliakbarpour for their advices and helps in the data acquisition.

References

Aggarwal, J., Cai, Q., Liao, W., Sabata, B., 1994. Articulated and elastic non-rigid motion: a review. In: Proceedings of the IEEE Workshop on Motion of Non-Rigid and Articulated Objects, pp. 2–14.

Badler, N., Phillips, C., Webber, B., 1993. Simulating Humans: Computer Graphics, Animation, and Control. Oxford Univ, Press.

Bartenieff, I., Lewis, D., 1980. Body Movement: Coping with the Environment. Gordon and Breach Science, New York, 1980.

Bobick, A., 1997. Movement, activity and action: the role of knowledge in the perception of motion. Philosophical Transactions of the Royal Society B: Biological Sciences 352, 1257–1265.

Cedras, C., Shah, M., 1995. Motion-based recognition: a survey. Image Vision Comput 13, 129–155.

Cheng, F., Christmas, W., Kittler, J., 2004. Periodic human motion description for sports video databases. In: Proceedings of the 17th International Conference on, Pattern Recognition (ICPR04).

Foroud, A., Whishaw, I., 2006. Changes in the kinematic structure and non-kinematic features of movements during skilled reaching after stroke: A laban movement analysis in two case studies. Journal of Neuroscience Methods 158, 137–149.

Gavrila, D.M., 1999. The visual analysis of human movement: A survey. Computer Vision and Image Understanding 73, 82–98.

Holte, M.B., Moeslund, T.B., Nikolaidis, N., Pitas, I., 2011. 3D Human action recognition for multi-view camera systems. In: International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), pp. 342–349, 16–19 May 2011.

Hutchinson, A., 1974. Labanotation. Oxford Univ, Press.

Khoshhal, K., Aliakbarpour, H., Quintas, J., Drews, P., Dias, J., 2010. Probabilistic LMA-based classification of human behaviour understanding using power spectrum technique, in: 13th International Conference on Information Fusion 2010. EICC Edinburgh, UK.

Khoshhal, K., Aliakbarpour, H., Quintas, J., Hofmann, M., Dias, J., 2011. Probabilistic LMA-based human motion analysis by conjugating frequency and spatial based features. In: 12th international Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS'11).

Khoshhal, K., Aliakbarpour, H., Mekhnacha, K., Ros, J., Quintas, J., Dias, J., 2011. LMA-based human behaviour analysis using HMM. In: L.M. Camarinha-Matos (Ed.), DoCEIS, vol. 349 of IFIP, Springer, pp. 189–196.

Khoshhal Roudposhti, K., Santos, L., Aliakbarpour, H., Dias, J., 2012. Parameterizing Interpersonal Behaviour with Laban Movement Analysis – A Bayesian Approach–. In: IEEE Computer Vision and Pattern Recognition Workshops'12 – SISM'12, June 2012.

Moeslund, T.B., Granum, E., 2001. A survey of computer vision-based human motion capture. Computer Vision and Image Understanding 81, 231–268.

Moeslund, T.B., Hilton, A., Kruger, V., 2006. A survey of advances in vision-based human motion capture and analysis. Computer Vision and Image Understanding, 90–126.

Murphy, K., 2005. Berkeley, Bayesnet toolbox. In: http://bnt.insa-rouen.fr/.

Pentland, A., Liu, A., 1995. Modeling and prediction of human behavior. IEEE Intelligent vehicles 95, 350–355.

Poppe, R., 2007. Vision-based human motion analysis: An overview. Computer Vision and Image Understanding 108, 4–18.

Poppe, Ronald A. A survey on vision-based human action recognition. Image Vision Comput. vol. 28, 6, (June 2010), pp. 976–990.

Ragheb, H., Velastin, S., Remagnino, P., Ellis, T., 2008. Human action recognition using robust power spectrum features, in: 15th IEEE International Conference on Image Processing. ICIP 2008, 753–756.

Rao, C., Yilmaz, A., Shah, M., 2002. View-invariant representation and recognition of actions. Internation Journal of Computer Vision, 203–226.

Rett, J., Dias, J., Ahuactzin, J.-M., 2008. Laban Movement Analysis using a Bayesian model and perspective projections, Brain, Vision and AI, 2008, iSBN: 978-953-7619-04-6.

Rett, J., 2008. Robot–human interface using laban movement analysis inside a bayesian framework, Ph.D. thesis, University of Coimbra, 2008.

Ryoo, M.S., Aggarwal, J.K., 2006. Recognition of composite human activities through context-free grammar based representation. In: CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, Washington, DC, USA, 2006, pp. 1709–1718.

Ryoo, M.S., Aggarwal, J.K., 2008. Recognition of high-level group activities based on activities of individual members. In: WMVC '08: Proceedings of the 2008 IEEE Workshop on Motion and video Computing, IEEE Computer Society, Washington, DC, USA, pp. 1–8.

Turaga, P., Veeraraghavan, A., Chellappa, R., 2008. Statistical analysis on Stiefel and Grassmann manifolds with applications in computer vision. In: IEEE Conference on Computer Vision and, Pattern Recognition, pp. 1–8, 23–28 June 2008.

Wang, L., Hu, W., Tan, T., 2003. Recent developments in human motion analysis. Pattern Recognition Society, 585–601.

Zhao, L., Badler, N., 2005. Acquiring and validating motion qualities from live limb gestures. Graphical Models, 1–16.