Probabilistic Human Activity Understanding

Exploring Relationship Between Human Body Motion and the Environmental Context

Kamrad Khoshhal, Jorge Dias

Institute of Systems and Robotics, Department of Electrical and Computer Engineering, University of Coimbra
{kamrad, jorge@isr.uc.pt}

Abstract. This paper presents an approach to model human activities based on relationship characteristics of body part motions and environment. Between human body motions properties and environment in each human activities are several relations that were less explored in the related works. To study human motion properties, Laban Movement Analysis (LMA), a known human motion descriptor was used. This work focused on *Relationship* component of the LMA to analyse human activities related to the environment. Bayesian network is a proper approach for the mentioned goal, in order to be able to predict 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. The results present the capability of the model to analyse any human activities related to environment using the LMA framework.

Keywords: Relationship characteristic, human activity understanding, human movement understanding, Bayesian approach, Laban movement analysis.

1 Introduction

This paper presents a new approach to analyse human activities based on the existed relations between different human body motions properties and environment parameters. Using some features of human body motions such as position, velocity, acceleration, trajectory, etc. have been used for analysing human motion and activities, but not only those features are important but also the relation between the features also is a very important characteristic of analysing different human movements. These relations can be defined in three levels. In the first level, the relations come from between different motion parameters of inside a body. Then the relation between one person and environment are important to obtain human behaviour in a scene. Finally human-human interaction can be analysed by defining the relationship of human motion characteristics of a couple of persons. For instance; in normal walking action, many features can be obtained, but between legs and hands there is a harmony which is very important to realize the action. In another example, clap ones hands together, the hands touching together is the relation of this two body parts. It can be defined also between a body and an external object such as reaching to or grasping a glass, and also other body such as hand shaking, punching, etc. In this work the two first levels were explored.

The idea comes from a known human motion analysis system, Laban Movement Analysis (LMA), which was created for choreography and dance notation purpose. The system has a framework that consists many types of features to interpret human motions. One of those types of features are provided a way to explain the relationship of a human body motion that can have with other objects (other body parts, environment, other body). Those types of relationship are infinite, and having a general structure that could be able to explain them as much as possible, is a big challenge that we attempt to formalize it by LMA framework.

Fig.1 presents a sequence of human movements, which can be analysed as a sequence of human interactions with a robot. These information can be useful in different application such as smart-home, surveillance, human robot interaction, etc.

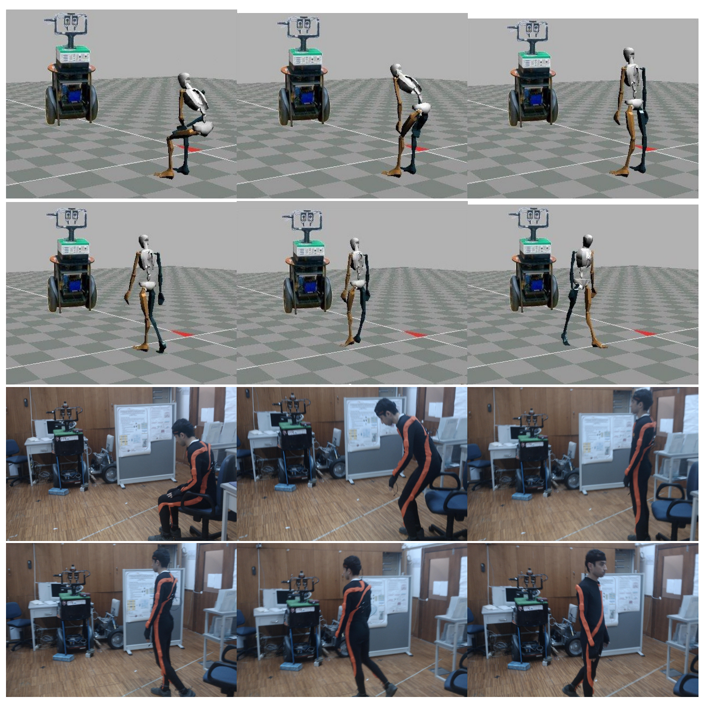

Fig. 1. Presenting a scenario that a human skeleton, obtained from a person using a suit motion tracker (MVN®), is doing some movements (rising, standing and walking) and interacting with a robot (reaching and spreading).

In our previous works, several features based on the LMA concepts, were defined to analyse human movements and activities. In [10] an approach was presented to use frequency-based features to estimate human movements based on some body parts acceleration signals. Then in [11], spatial-based feature also was used to improve the previous estimation results. All those work was based on 3D data 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 features. To implement the idea, a Bayesian Network (BN) which is a popular approach in this area, was used. The capability of generalization of the system because of using LMA framework is the one of the big advantageous of this approach.

1.1 Previous related works

Analysing human motion is a prerequisite for understanding any human activities, such as human behaviour, human-robot or human-human interaction, etc. Analysis of human activities can be investigated in different levels. Bobick in [5] 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 or previous knowledge to be identified, but to understand human action or behaviour we need to have a sequence of human movements or states, related to the environment or scenario.

There are many different kinds of human movements that researchers try to recognize. Each of them can be identified with 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 good behaviour analysis. Legs, hands, head and face are the most important body parts to track in different related applications. Some examples can be given such as in gait recognition, where the dynamics of leg motion is important; for facial expression recognition the focus lies solely on the face while in motion recognition, all parts of body can be involved, where hands and head stand out as the most important ones.

There are many surveys about human motion analysis such as Aggarwal et al. in [1] which covered various methods used in articulated and elastic non-rigid motion. Cedars and Shah in [6] 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 [8] described a work in human motion analysis in terms of 2D and 3D approaches. Pentland in [15] touched on several interesting topics in human motion analysis and its applications. Moeslund and Granum [13,12] presented a survey of computer-vision-based human motion capture problems (initialization, tracking, pose estimation and recognition). Wang et al. in [22] provided a comprehensive survey of research on computer-vision-based human motion analysis (human detection, tracking and activity understanding). Poppe in [16] described the characteristics of human motion analysis, and the study divided the analysis into modeling and estimation phases.

As Bobick [5] said to analyse human activities, we need to know the underlying movement concept. It means to reach to human activities, we need to find the relation of between human movements and environment parameters. There are infinite relation that can be seen in human activities, Thus researchers always define some specific relationship properties to present their methods. For instance; Rao et al in [17] 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 like 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 to have the mid-level definition of human motions. To have this mid-level, several works have been done for several type of features based on the LMA systems [7,11,10,19,23], all of them were not explored in relationship component parameters of LMA, but in *Effort*, *Shape* and *Space*.

For implementing the idea, Bayesian Networks (BN) has been used. BN presents many advantages on using prior knowledge and modeling the dynamic dependencies between parameters of object states. In related fields, this approach is popular and researchers have been keen on applying it. One example is Rett [18] who applied a general BN framework for analyzing human motions based on LMA concept for a human machine interaction application, however in that work the relationship parameters were not used. Ryoo and Aggarwal [20,21] presented a framework for human action modeling by using a number of BNs to recognize the poses of body parts. There are several reasons to use a Bayesian approach in our application: 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 will play the relationship roles, but it's very depend on the learning process. The point is, the relations parameters in the model were not defined explicitly, and it is sensitive to the learning process. By considering those relationships also as some parameters of our model, more reliable results will be achieved. Finally by connecting human body motions to the scene, which is one of the Relationship purposes, a probabilistic model for analysing human activities was performed.

2 Features Extraction

Several features from human motion can be extracted in different domains. In our previous works, frequency-based features [10] and spatial-based feature [11] were extracted to analyse several general human movements. In this paper, we intent to use those features which rely on relation between human motions and environment. Based on the relationship definitions which will be presented in next section, infinite features can be defined. To perform an experiment and depend on the activities, it is needed to collect the best features which can distinguish the activities perfectly. In this work, two types of features were extracted as explained in next subsections.

2.1 Body motion based features:

In this step, we intend to estimate some general human activities, like Walking, Running, siting, rising, falling down, standing, etc. without using environmental parameters. Thus, we attempt to define the relationship parameters between body parts of one person.

In walking types activities, usually there is a harmonic motion in hands and feet related to the body center. Thus we use acceleration signals of human body parts related to the body center, and then Fast Fourier Transform (FFT) and Power Spectrum (PS) signals of the acceleration signals were extracted [10] (see Fig.2). The peaks of PS signals (coefficients) in four frequency domains (1-10 HZ, 11-20 HZ, 21-30 HZ, 31-40 HZ) are collected as the low level features (LLFs).

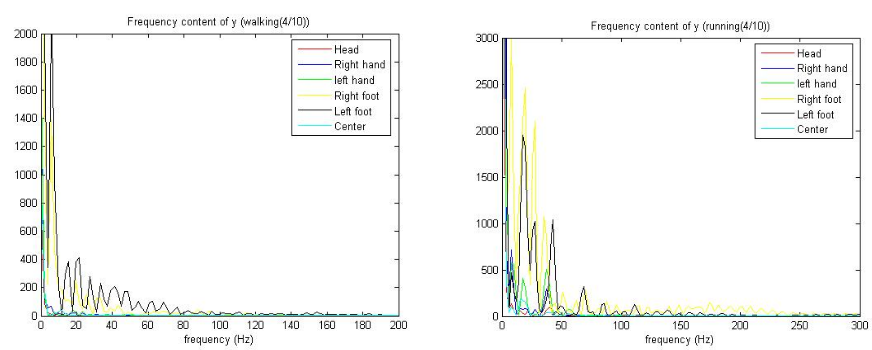

Fig. 2. PS results of some body parts acceleration signals for (left) walking and (right) running movements

Another useful features to distinguish the activities are the relation of two objects like head and feet in terms of the relation between those body parts during different movements. For instance; in sitting and falling down movements, distance of those body parts reduce and in a rising movement there is a opposite situation, and in standing or walking there is no considerable changing in that aspect. Thus the difference of distance between head and feet during a movement can be obtained:

$$\Delta D = \sum_{t=2}^{n} \left(\left(X_{obj1}^{t} + X_{obj2}^{t} \right) - \left(X_{obj1}^{t-1} + X_{obj2}^{t-1} \right) \right), \tag{1}$$

where X_{obj1}^{t} and X_{obj2}^{t} denote position of the two objects $obj1$ and $obj2$ at time t, respectively. n denotes the number of frames inside of the window signal, and ΔD denotes the difference of the two objects distances between the first and end of the window signal in the meter unit.

2.2 Environmental based features:

To implement this part, a scenario needed to be defined to present the idea more visible. Thus a scenario which includes a static robot agent that people can interact with that, was defined. Some relevant activities were performed like, *reaching*, *spreading*, and *passing*. The input features in this level will be the probability of human movements and the difference of distance during human activities with an object like a robot using equation 1.

3 LMA Relationship Definition

Laban Movement Analysis (LMA) is a well-known method for observing, describing, notating, and interpreting human movement [4,3]. 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 [3,23] 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 happening (for instance; if it is sudden or sustained), *Shape* has another view, and explains human body shape (as a blob) deformation during a movement. Several works were published around these components, but there is no any attempt to analyse human activities in relationship's component aspect which explain human body parts motion relations to other parts of the body, another body or environment parameters [19].

In the *Relationship* component of LMA has several parameters that categorized different types of possible relations. As Huchinson [9] mentioned those parameters were listed and described as follows; 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 this application to analyse a person movements and activities (interact with a static object (see Fig.1), such as reaching, passing and spreading), some of the parameters which are more representative were used to present the approach. The approach allows us to generalize the system. The used relationship parameters are described as follows:

- *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 approach to or move away from your 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 retain for more than the moment usually depend 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.

3.1 ***Relationship*'s component modelling**

As can be realized from the parameters definitions, there are several properties to analyse any human activities. To simplify the system, some of the properties were modeled for the performed activities.

- *passing* and *retention* relationship:

These properties can be used between every two objects. In this model it was applied between two body parts. During each human activities, there are different motion signals between body parts. For instance, in a walking type movement, there is a same motion signal between opposite sides of hands and feet in the same moment. To analyse those signals, characteristics of passing and retention relationship were used between each body parts related to a reference like body center (if a body part is in *passing* or *retention* relationship with the body center). If the frequency-based quantities were extracted for each body parts movements is more than a threshold, it means that there is a passing motion depend on the body center, and it states *passing* relationship, otherwise states *retention* one. Based on the training dataset, the frequency-based quantities were discretized by these thresholds:

$$f_{pb}^{i} = \begin{cases} No & Max\{f_{pb}^{i}\} <= 20 \\ Low & 20 < Max\{f_{pb}^{i}\} <= 150 \\ Medium & 150 < Max\{f_{pb}^{i}\} <= 1000 \\ High & Max\{f_{pb}^{i}\} > 1000 \end{cases}$$

where $Max\left\{f^i_{pb}\right\}$ denotes a frequency-based coefficient in i^{th} frequency sub-domain for a pb body part. Thus for each body part, one variable with a double state is defined:

$$Pass_{pb}\epsilon\left\{passing, retention\right\} \tag{2}$$

– *Toward* and *Away* relationship:

They can be used in two different spaces; between body parts and between a person and another object. By having the Equation 1's output, between two objects , *Toward* qualities can be estimated by discretizing of the $\varDelta D$:

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

where P, S and N denote positive, Still and Negative qualities of *Toward* property, respectively. These property were used between head and feet as two body parts of one person ($\varDelta D_b$), and a person and a robot as two objects in a scene ($\varDelta D_e$) (see Fig.3). Thus there are two Toward variables with the three states.

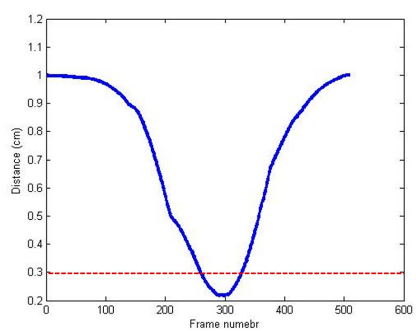

Fig. 3. The difference of distance between a person and a robot, while the person is walking and passing near of the robot. The red dashed line presents the defined threshold.

By combining The *Toward* variable which defined between two body parts and *pass* variable, the human movements were estimated.

$$Movement\epsilon\left\{walking, running, sitting, rising, falling, standing\right\} \tag{4}$$

And Finally using The Toward variable between person and robot and the Movement variable, human activities can be estimated. the activity states are:

$$Activity \epsilon \{reaching, spreading, passing, other\} \tag{5}$$

Toward property between human and robot is the proper feature to analyse *reaching* and *spreading* activities. And *passing* activity can happen when the *Toward* variable states Still state while the person is walking. Thus there are more possibilities to define more activities by having those movement and the environmental parameter states. And it was not supposed to implement a complex model, but to present an approach which can be easily generalized the idea based on that.

4 Bayesian Model

Bayesian approach is a popular method to classify human motions and activities [21,19]. 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.

In the model (see Fig.4), 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) were used for the *Passing/retention* relationship of the same body part:

$$P\left(Pass_{pb} \mid \prod_{i=1:4} Max\left\{f_{pb}^{i}\right\}\right) = \frac{P\left(Pass_{pb}\right) \prod_{i=1:4} P\left(Max\left\{f_{pb}^{i}\right\} \mid Pass_{pb}\right)}{\prod_{i=1:4} P\left(Max\left\{f_{pb}^{i}\right\}\right)} \tag{6}$$

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

$$P\left(Toward_{b} \mid \Delta D_{b}\right) = \frac{P\left(Toward_{b}\right) P\left(\Delta D_{b} \mid Toward_{b}\right)}{P\left(\Delta D_{b}\right)} \tag{7}$$

and the probability of similar property but between a person and an external object can be achieved by:

$$P\left(Toward_{e} \mid \Delta D_{e}\right) = \frac{P\left(Toward_{e}\right) P\left(\Delta D_{e} \mid Toward_{e}\right)}{P\left(\Delta D_{e}\right)} \tag{8}$$

There are two other levels, that one of them is movement's level which is not related to the environment parameters. And the other level which is named activity level, to connect human movement information to the scene in order to reach human activities. Those two probabilities models are:

$$P\left(M \mid Toward_{b} \prod_{pb=1:n} Pass_{pb}\right) = \frac{P\left(M\right)\ P(Toward_{b}|M) \prod_{pb=1:n} P\left(Pass_{pb}|M\right)}{P\left(Toward_{b}\right) \prod_{pb=1:n} P\left(Pass_{pb}\right)} \tag{9}$$

$$P\left(Act \mid Toward_e\, M\right) = \frac{P\left(Act\right)\, P(Toward_e|Act)\, P\left(M|Act\right)}{P\left(Toward_e\right) P\left(M\right)} \tag{10}$$

where M and Act denote movement and activity, respectively, and n denotes the number of used body parts.

Figure 4 presents the dependencies of those all different variables (LLFs, *Relationship*'s component, movement, activity) in two different space (related or not related to the environmental parameters) in one model.

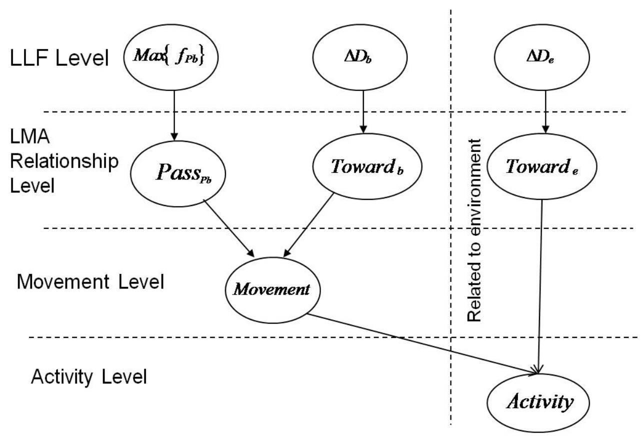

Fig. 4. Global Bayesian model

5 Experiments

To obtain the input data (3D position and acceleration of body parts) a body motion tracker (MVN®suit) [11] was used. The suit used a global reference in the scene as a start point which is known. The distance between human body and the robot during the activities can be calculated. Another approach to obtain the data, is using a network camera framework using Aliakbarpour's method [2], which gives the human body position in the scene. Several different human movements were performed. One second's window which shifts half second during each record, was defined to process the all performed movements and activities. Two series of data were achieved; human body parts positions related to the body center, and the person 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 [14] was used to implement the 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 which are the input of activity level of model. Those data were obtained from around 100 sequence of human movements with different durations.

	Walking	Running	Sitting	Rising	Falling	Standing	%
Walking	63					2	96.92
Running	1	72			2		96.00
Sitting			46		2		95.83
Rising				34		1	97.14
Falling		1		1	26		92.85
Standing						155	100

Table 1. Human movement classification result

Fig.5 presents the results of the activity classifier in top view, and Fig.6 presents probability histograms of the both classifiers of movement and activity in three middle steps. In these samples, in the left image, the person is walking and passing near of the robot, and in the right image, however same activities are performed but the person is in running movement state.

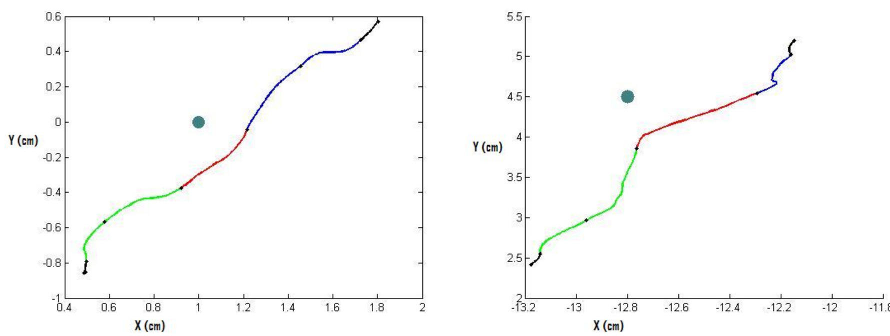

Fig. 5. Two samples of the activity classification results (top view). The green, red, blue and black lines present trajectory of a person with *Reaching*, *Passing*, *Spreading* and *Other* states, respectively. The circle presents the position of the robot in the scene (left: walking movement and right: running movement).

Table.2 presents the result of the model for human activities level. The result shows that our model for passing activity is not accurate as others. The reason is that the related features is not sufficient enough for all types of passing activity, such as running situations.

	Reaching	Spreading	Passing	Other	%
Reaching	138		1	7	94.52
Spreading		149	1	5	96.13
Passing	9	2	45	3	76.17
Other	1	1	4	144	96.00

Table 2. Human activity classification result

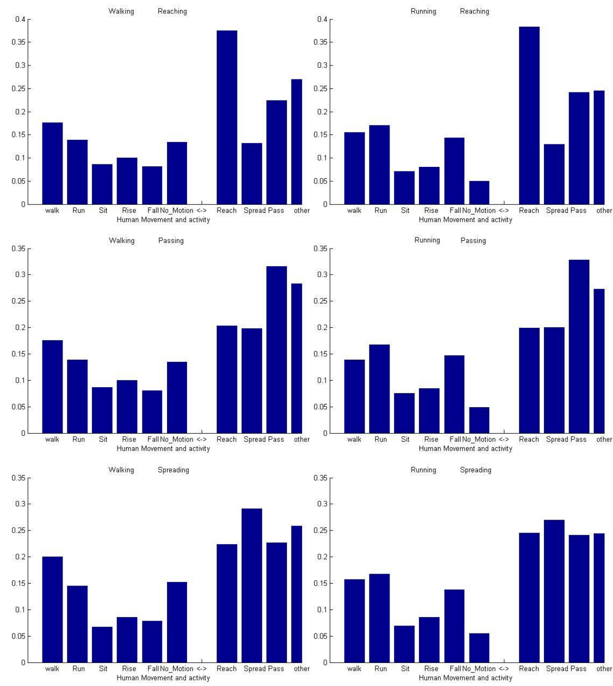

Fig. 6. A Triple step of changing the human activity states from two sequence of human movements which were presented in Fig.5 in top view also, Left: the person is walking. Right: the person is running.

By having the obtained information by the model, 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), fast accident (reaching in running or falling down movement's states), etc. It means, these relationship's parameters can assist us to analyse even very 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 human activities through the relationship concept. There are infinite relations between body part motions and environment in human activities. Using *Relationship* component of LMA is the key of the system which allows us to analyse any human activities specially to interact with environment. In this work, we attempted to computerize the component for some performed activities. A Bayesian Network is defined to develop a model which can fuse different types of features, and classify the movement and activities. In the future, we intend to investigate *Relationship* component in interpersonal activities.

7 Acknowledgment

This work was financially supported by a PhD grant (SFRH/BD/70640/2010) from the Portuguese Foundation for Science and Technology (FCT).

References

1. J.K. Aggarwal, Q. Cai, W. Liao, and B. Sabata. Articulated and elastic non-rigid motion: a review. pages 2–14, 1994. Proceedings of the IEEE Workshop on Motion of Non-Rigid and Articulated Objects.
2. H. Aliakbarpour, J. F. Ferreira, K. Khoshhal, and J. Dias. A novel framework for data registration and data fusion in presence of multi-modal sensors. In *Emerging Trends in Technological Innovation*, volume 314/2010 of *IFIP Advances in Information and Communication Technology*, pages 308–315. Springer Boston, 2010 edition, February 2010.
3. N.I. Badler, C.B. Phillips, and B.L. Webber. *Simulating Humans: Computer Graphics,Animation, and Control*. Oxford Univ. Press, 1993.
4. D. Bartenieff, I. & Lewis. *Body Movement: Coping with the Environment*. Gordon and Breach Science, New York, 1980.
5. A.F. Bobick. Movement, activity and action: the role of knowledge in the perception of motion. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 352(1358):1257–1265, 1997.
6. C. Cedras and M. Shah. Motion-based recognition: a survey. *Image Vision Comput*, 13:129–155, 1995.
7. 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. *Journal of Neuroscience Methods*, 158:137–149, 2006.
8. D. M. Gavrila. The visual analysis of human movement: A survey. *Computer Vision and Image Understanding*, 73:82–98, 1999.
9. Ann Hutchinson. *Labanotation*. Oxford Univ. Press, 1974.
10. Kamrad Khoshhal, Hadi Aliakbarpour, Joao Quintas, Paulo Drews, and Jorge Dias. Probabilistic lma-based classification of human behaviour understanding using power spectrum technique. In *13th International Conference on Information Fusion2010*, EICC Edinburgh, UK, July 2010.
11. Kamrad Khoshhal, Hadi Aliakbarpour, Joao Quintas, Martin Hofmann, and Jorge Dias. 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)*, 2011.
12. Thomas B. Moeslund, Adrian Hilton b, and Volker KruÂšger. A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, pages 90–126, 2006.
13. Thomas B. Moeslund and Erik Granum. A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding*, 81:231–268, 2001.
14. Kervin Murphy and Berkeley. Bayesnet toolbox. In *http://bnt.insa-rouen.fr/*, 2005.
15. Alex Pentland and Andrew Liu. Modeling and prediction of human behavior. *IEEE Intelligent vehicles 95*, pages 350–355, 1995.
16. Ronald Poppe. Vision-based human motion analysis: An overview. *Computer Vision and Image Understanding*, 108:4–18, 2007.

17. C. Rao, A. Yilmaz, and M. Shah. View-invariant representation and recognition of actions. *Internation Journal of Computer Vision*, pages 203–226, 2002.
18. Joerg Rett. *ROBOT-HUMAN Interface using LABAN Movement Analysis Inside a Bayesian framework*. PhD thesis, University of Coimbra, 2008.
19. Joerg Rett, Jorge Dias, and Juan-Manuel Ahuactzin. *Laban Movement Analysis using a Bayesian model and perspective projections*. Brain, Vision and AI, 2008. ISBN: 978-953-7619-04-6.
20. M. S. Ryoo and J. K. Aggarwal. 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*, pages 1709–1718, Washington, DC, USA, 2006. IEEE Computer Society.
21. M. S. Ryoo and J. K. Aggarwal. 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*, pages 1–8, Washington, DC, USA, 2008. IEEE Computer Society.
22. Liang Wang, Weiming Hu, and Tieniu Tan. Recent developments in human motion analysis. *Pattern Recognition Society*, pages 585–601, 2003.
23. L. Zhao and N.I. Badler. Acquiring and validating motion qualities from live limb gestures. *Graphical Models*, pages 1–16, 2005.