# Improved Semantic-based Human Interaction Understanding Using Context-based knowledge

Kamrad Khoshhal Roudposhti *, Jorge Dias *,**
* Institute of Systems and Robotics, Department of Electrical and Computer Engineering, University of Coimbra, Portugal.
** Khalifa University, UAE
{kamrad,jorge@isr.uc.pt}

***Abstract*—This paper proposes a descriptive approach for context-based human activity analysis through an hierarchical framework in a scene understanding application. Each human movement with respect to himself, others and scene, can arise different layers of human activities analysis, which usually investigated separately depend on the application. Human behaviour can not be analysed properly, since the all different layers of information were not considered. The effect of using the different layers of information to increase the accuracy of the analysis is presented in the study. The contributions are, using different information layers such as human body parts movement and human-object interaction, in 3D space, to improve human activity analysis, and proposing a probabilistic and descriptive model, based on a well-known human movement descriptor and Bayesian Network (BN) approach. Thus, based on the mentioned framework, the model is generalizable and flexible which are necessary for having such an applicable system. The capability of the proposed approach is presented in the experiment's section.**

***Index Terms*—Scene understanding, hierarchical framework, human interaction analysis,Bayesian approach, human movement analysis, descriptive model.**

## I. Introduction

This paper proposes a flexible scene understanding model, which can describe human activity based on a well-known descriptor, and deal with uncertainly using probabilistic models. Human activity analysis can be categorized as context-free and context-based. In context-free based approaches the model is independent of scene parameters, and just rely on the features belong to the person. However in the reality, context-based features play very important role to analyse human activities. For instance, when a person going to reach a chair, we will realize that properly the person going to sit on the chair, not to sleep.

As Delaitre et al, described in [6], since object detection is a widely studied topic in computer vision, analysing the relation between human movements and the existent object around, can produce valuable information for human daily activities. For instance, people have been learned the (most probable) normal activities when the person is reaching to a chair, thus people have a probabilities set of activities depend on the objects in the scene.

The problems is, what level of human movements information might be useful, and then how a general framework can be defined for analysing any possibility of human-object interactions. For the mentioned aspect, from the low level information such as body parts motions to higher ones such as human interactions can be useful. Dealing with the mentioned different information caused a complex model. Thus, an hierarchical framework was used to reduce the complexity of the model [1] to provide different level of human activity analysis [11].

The relationship probability distributions between human motions and human-object based information, can be modelled, by given the possible activities and the interested objects in a scene. Laban Movement Analysis (LMA) system which consists of several components, is used to define proper human motions (*Effort*, *Shape*) [13], [12]and human-scene relations (*Relationship*) [16], [10] variables. Gupta et al. in [9] tackled the problem based on the 2D images. Thus they focused more on the computer vision problems for the mentioned applications, and just used the person hand trajectory information to analyse human-object interactions (reaching and manipulation). Their mentioned Bayesian model can not deal easily with the extension of the work for other activities. Thus we proposed the hierarchical model to deal with the problem, and to avoid the limitation of the 2D-based analysis, we used a motion tracker suit (MVN®) with several inertial sensor attached on the different body parts to have 3D pose of human body parts with maximum 120 frames per second resolution. However there are several works using 3D-based human movement analysis with high accuracy [14], [4], and also in 3D virtual applications [7], but only focused on classifying simple human movements.

This paper is organized as following; Sec. II presents the feature extraction methods, and then based on that, the hierarchy-based human activity modelling is presented in Sec. III. Experimental results presented and discussed in Sec. IV, and Sec. V closes the paper with a conclusion and an outlook for future works.

## II. Feature Categorization and Extraction Using LMA

Body parts trajectories during human activities and the relationship between human and interested objects in the scene, are the input data of this study. A motion tracker suit is used to obtain the 3D human body parts positions

978-1-4799-0652-9/13 $31.00 © 2013 IEEE
DOI 10.1109/SMC.2013.494

with respect to a global reference in the scene. Then other information can be extracted easily for each body parts such as velocity, acceleration, angular acceleration, angular velocity, etc. The problem is that which one of those features are needed and how we can used them to model any human movement-based activities. For analysing human activities inside a scene, the needed features can be categorized in a couple of parts; body parts motions-based features for human movement analysis, and human-object relationship-based features for human-object interaction analysis.

The mentioned categories can be realized from a well-known human movement descriptor, Laban Movement Analysis (LMA), in choreography science, which the needed features were defined, semantically [3]. LMA consists of several components, four of them (*Effort, Space, Shape, Body*) defined for body motion analysis [18], [13], and one of them (*Relationship*) for existent relations between human and objects inside a scene [16], [10].

Effort component describes that how a performer consume his/her energy during his/her movements, by some sub-components that each of them has a couple of bipolar states. For instance, *Effort.time*'s sub-component explain that if a body part moves suddenly or sustainedly, during an specific movement. The mentioned components were modelled and implemented probabilistically [12], [15].

In [13], frequency-based features from five body parts acceleration signals, were extracted to analyse *Effort.time's* property. As can be seen in Figure 1, Power Spectrum (PS) signals of the obtained 3D acceleration signals were estimated. The signals show the difference between two types of human movements which are similar in spatial domain, very explicitly. Khoshhal and Dias proved that collecting first four coefficients of the PS signals are sufficient for estimating the *Effort.time* sub-component, by using Equation 1, which presents the Bayesian model [13]. As mentioned before *Effort.time* has a couple of states; *sustained*, *sudden*.

$$P\left(E_a^{bp} \mid \prod_{i=1:4} Max\left\{{}^i f_a^{bp}\right\}\right) = \frac{P\left(E_a^{bp}\right)\prod_{i=1:4} P\left(Max\left\{{}^i f_a^{bp}\right\} \mid E_a^{bp}\right)}{\prod_{i=1:4} P\left(Max\left\{{}^i f_a^{bp}\right\}\right)} \quad (1)$$

where $P\left(Max\left\{{}^i f_a^{bp}\right\}\right)$,$P\left(E_a^{bp}\right)$ denote probability of $i$th coefficient's PS signal of $bp$'s body part acceleration of person $a$, and probability of *Effort.time* sub-component ($E$) of $bp$'s body part of person $a$, respectively.

*Shape* and *Space,* explain about the spatial-based characteristics of human movements. Shape focus on the deformation of human body as a blob in three planes during a movement by some states, for instance; sinking or rising (in vertical plane), retreating or advancing (in sagittal plane), and enclosing or spreading (in horizontal plane) [13]. Space focus on the direction of body parts motions (trajectory) in the vertical, horizontal and sagittal planes [18], [15].

*Shape* component in vertical plane was used in this study as spatial-based features. The related states are *sinking* and *rising*, which obtained by difference distance between head and feet during the window slide of human movements. Equation 2 present the model of *Shape.vertical* property.

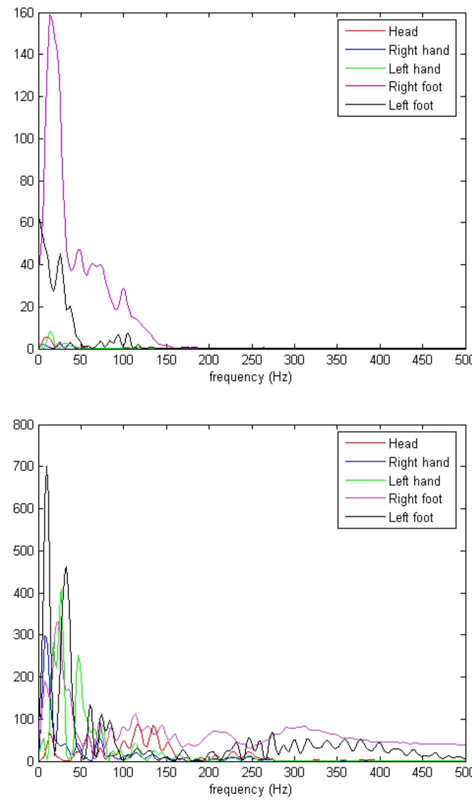



Figure 1. Power spectrum signals extracted from body parts acceleration signals through a walking (top) and a running (down) movement [13].

$$P\left(S_a^v \mid \Delta D_b\right) = \frac{P\left(S_a^v\right) P\left(\Delta D_b \mid S_a^v\right)}{P\left(\Delta D_b\right)} \quad (2)$$

where $P\left(S_a^v\right)$, $P\left(\Delta D_b\right)$ denote probability of *Shape* component in vertical plane for person $a$, and probability of human-height's change of person $a$, respectively. The difference distance was discretized in triple states by obtained thresholds in [13]; Negative (Decreasing distance), Positive (increasing distance), and Still (no significant difference).

Relationship is the only component of the LMA, which is defined for the existent connections between human movements and the scene [10]. It categorized in several types of possible relations between human body parts movements and scene (himself, object and other person). Khoshhal and Dias in [16] attempted to model some of the mentioned relationships (*Toward/Away*, *Contact*) during some human interactions (reaching, spreading, passing, following, handshaking, pushing, Non). Toward/Away's variable describes if one object like a person going to reach to other object like other person or a chair. *Contact*'s variable explain if one object touched to other object or not.

Based on [16] the mentioned Relationship properties were modeled as can be seen;

$$P\left(R_{o1-o2} \mid \Delta D_e\right) = \frac{P\left(R_{o1-o2}\right) P\left(\Delta D_e \mid R_{o1-o2}\right)}{P\left(\Delta D_e\right)} \quad (3)$$

| Category | LMA Component | Variable | States | Feature's Domain |
|---|---|---|---|---|
| Body parts motion | Effort | Time | Sudden / | Frequency |
| | | | Sustained | |
| | Shape | Vertical | Rising/ | Spatial |
| | | | Still/ | |
| | | | Sinking | |
| H-O relationship | Relationship | Toward (H-O1) | Toward/ | Spatial |
| | | | Still/ | |
| | | | Away | |
| | | Toward (H-O2) | Toward/ | |
| | | | Still/ | |
| | | | Away | |
| | | Contact (H-O1) | Connected/ Disconnected | |

Table I
DIFFERENT LMA COMPONENTS IN A COUPLE OF CATEGORIES WHICH ARE DEFINED FOR HUMAN INTERACTION ANALYSIS. H-O1 DENOTES HUMAN-OBJECT1. IN THE EXPERIMENT, O2 IS USED AS ANOTHER PERSON, AND O1 AS A CHAIR.

where $\Delta D_e$, $R_{o1-o2}$ denote the difference distance between two object which both don't belong to one person, and a *Relationship* property between two objects $o1$ and $o2$ such as (a person and a chair, and two hands of two people), respectively.

Depend on the scenarios and input data, some of the mentioned components are useful as can be seen on others previous works ([5], [8], [17], [18], [15]). Table.I presents all defined LMA parameters based on the three components for this study. However the model is extend-able for the other components.

## III. HIERARCHY-BASED HUMAN ACTIVITY ANALYSIS

Human movement can be analysed by having body parts motion information. The features of human motion were extracted in both temporal and spatial domains using the LMA structure [13]. The LMA.*Effort* component deals with temporal domain and the LMA.*Shape* component with spatial ones. Table I shows the mentioned features with the related states in the body part motions category.

*Effort.Time* variable is applied for each body parts during different movements (*Walking, Running, Falling down, Sitting, Rising, Standing*). For instance, in *Running's* movement usually there are higher probability of the *sudden* state for feet than in *Walking's* movement. For discriminating some of movements like *Rising* and *Sitting*, spatial-based features are better representative property[13]. Thus *Shape.Vertical* variable is used to deal with the spatial domain's property.

For analysing human-object interactions we used LMA.*Relationship* component which can be seen on the Table I in human-object relationship category.

Based on the mentioned features, several human-object interactions can be analysed. For this study, depend on the interested objects, a set of interactions was defined. Sitting and standing up actions for human-chair interaction, pushing and handshaking actions for human-human interaction's purpose, and reaching and spreading in general were defined. Table.II shows the defined classes for both movement and interaction variables.

| Variable | Classes |
|---|---|
| Movement | *Walking, Running, Falling down, Sitting, Rising, Standing* |
| Interaction | *Reaching, Spreading, Sitting on the chair, Standing up, Handshaking, Pushing, Other* |

Table II
HUMAN MOVEMENT AND INTERACTION CLASSES. *Other* MEANS ANY ACTIVITIES WHICH ARE NOT BELONG TO THE DEFINED CLASSES.

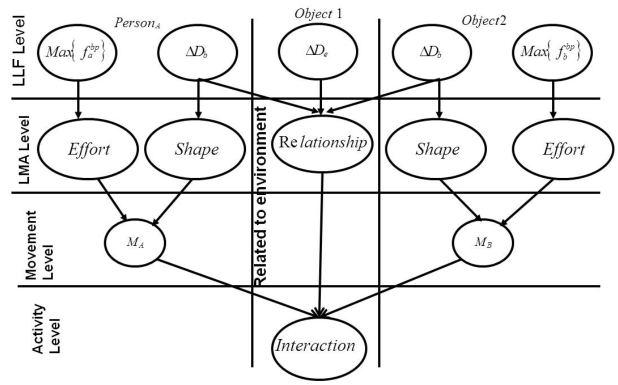



Figure 2. The hierarchical framework for scene understanding. LLF's level contains frequency-based features ($Max\left\{f^{bp}\right\}$) and spatial-based ones ($\Delta D_b$ for body parts motion's category and $\Delta D_e$ for human-object relationship). LMA's Level contains *Effort* and *Shape* components of the people and *Relationship* component. In the Movement level, we have $M_a$ and $M_b$ which denote movement class belongs to person $a$ and $b$ given their related *Effort* and *Shape* components, respectively. Finally Interaction class are estimated given both person $a$ and $b$ movement classes and the existent *Relationship* states.

### A. Bayesian Network Modeling

Bayesian Network (BN) is a well-known approach to model an hierarchical-based analysis [16], because of its flexibility and capability of fusion different types of features, and deal with uncertainly, decision making problem and prediction process. The Bayesian graphical model for the mentioned system can be seen in Figure 2 which presents the dependencies between the different levels information.

In each level of the BN, the probability of defined variables are modeled by Bayesian rule formulation. In the highest level, we intend to estimate the probability of each human interactions states, given the movement states probabilities of both person $a$ and person $o1$, and the relation between person $a$ and the two defined objects ($o1$ which is other person and $o2$ which is a chair). Thus the Equation 4 presents the mentioned Bayesian rule, which can see the dependencies.

$$P\left(I_a|M_a, M_{o1}, R_{a-o1}, R_{a-o2}\right) = \frac{P(I_a)P(M_a|I_a)P(M_{o1}|I_a)P(R_{a-o1}|I_a)P(R_{a-o2}|I_a)}{P(M_a)P(M_{o1})P(R_{a-o1})P(R_{a-o2})} \tag{4}$$

where, $I_a$,$M_a$ and $R_{a-o1}$denote person $a$'s *Interaction*, *Movement* and *Relationship* with respect to object $o1$, variables respectively.$P\left(M_a|I_a\right)$ denotes the estimation of *Movement*'s states of person $a$ probabilities given probability of its $I_a$ states.

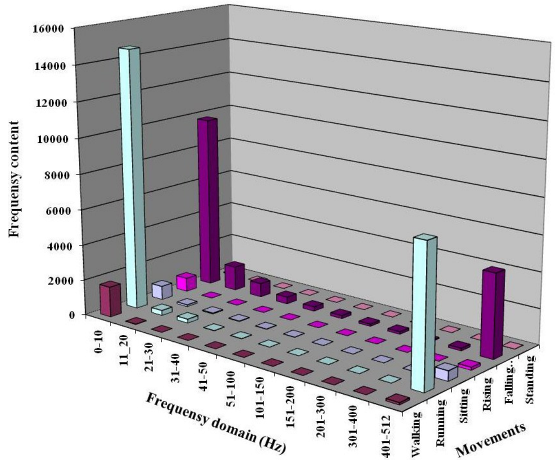



Figure 3. Histogram of the frequency-based features of a specific body part for different movements in different frequency sub-domains [16].

In the Equation 4 some variables which are located in lower levels, need to be solved. For instance; Equation 5 which proposed in [13], was used to model human movement.

$$P\left(M_a|E_a^{bp}S_a^v\right) = \frac{P\left(M_a\right)P\left(E_a^{bp}|M_a\right)P\left(S_a^v|M_a\right)}{P\left(E_a^{bp}\right)P\left(S_a^v\right)} \quad (5)$$

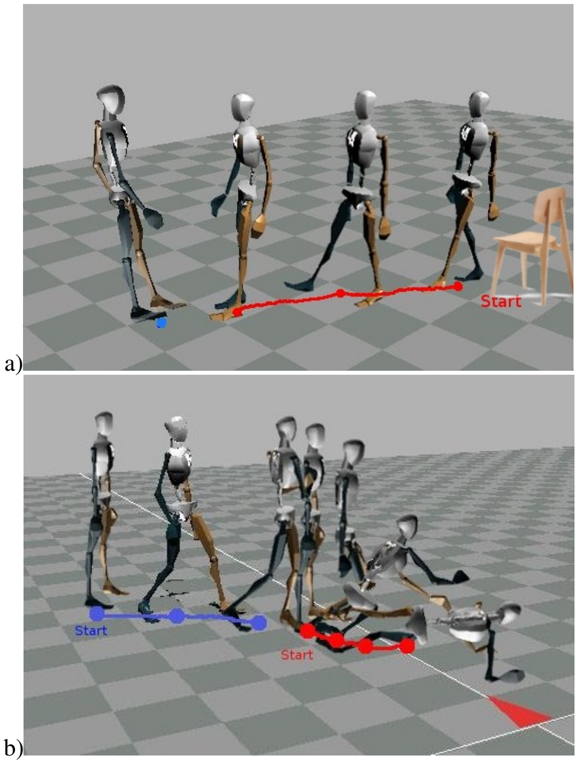
where $E_a^{bp}$and $S_a^v$ denote *Effort* component of LMA for $bp$'s body part of person $a$, and *Shape* component of LMA for person $a$ in vertical plane, respectively. $bp$ is the index of body parts which are used (hands, feet and head). $P\left(E_a^{bp}\right)$ , $P\left(S_a^v\right)$and $P\left(R_{a-o1}\right)$ denote probability of *Effort*, *Shape* and *Relationship* components of LMA, respectively. The mentioned LMA components probabilities are estimated by [13], [16], given frequency and spatial based features.

### B. Learning process

To obtain conditional probability of each variable in different levels of analysis, learning process is needed. Maximum likelihood is a well-known approach in the learning process, however there are several approaches which can be used for learning process [**?**]. For each class of movement, several data by the motion tracker suit were collected. The interested features were extracted for different body parts (feet, hands and head). To analyse the conditional probability for a variable in each level, histogram-based approach was used.Figure 3, presents a sample of the obtained histogram of the frequency content in different sub-domains for different types of human movements [16].

Then after applying learning process in the all levels, we will be able to obtain the $P\left(Max\left\{{}^i f_a^{bp}\right\} \mid E_a^{bp}\right)$ to estimate Equation 1, $P\left(\Delta D_b \mid S_a^v\right)$ to estimate Equation 2, $P\left(\Delta D_e \mid R_{o1-o2}\right)$ to estimate Equation 3, $P\left(E_a^{bp}|M_a\right)$ and $P\left(S_a^v|M_a\right)$ to estimate Equation 5, and $P\left(M_a|I_a\right)$, $P\left(M_{o1}|I_a\right)$, $P\left(R_{a-o1}|I_a\right)$ and $P\left(R_{a-o2}|I_a\right)$ to estimate Equation 4.

a)




b)

Figure 4. The scene with a couple of samples of people activities, provided by the motion tracker suit (MVN®)

## IV. EXPERIMENTS

For collecting data, a motion tracker suit, which gives body parts 3D position, is used. A person wore the suit and after doing the calibration process (for suit's sensors), the interested activities were performed, and the outputs of the suit as a XML file for all record series were stored. The data collected by two different people. Each person in different acquisition times, has needed to perform calibration process, and it means for an specific person, the data could be variant in different times of the calibration process.

The motion tracker gives the body parts positions depend on a global reference, which is defined in the calibration step. Based on that global reference the others interested objects positions in the scene, were defined (see Figure 4). For feature extraction step the sliding window approach was selected. The window size has defined one second, and shifting the window by half of the window size. The frame rate of the system is 120 f/s. Ten different sequences (each sequence contain more than 1000 frames) were collected for each type of human movements, which performed inside of the different actions and interactions.

As mentioned before, frequency-based features are estimated by using PS signals (Figure. 1) which obtained by FFT of human body part's acceleration signals (as can be seen the detail on [13]). Spatial features were extracted by difference distance of the interested objects during the selected window slide (more details can be seen on [16]). For Learning

| Frame | Level | States with their probability |
|---|---|---|
| 180 | | |
| | LMA.Eff.T-Head | Sudden:45%, Sustained:**55%** |
| | LMA.Eff.T-LFoot | Sudden:**63%**, Sustained:37% |
| | LMA.Eff.T-RFoot | Sudden:**61%**, Sustained:39% |
| | LMA.Eff.T-RHand | Sudden:**55%**, Sustained:45% |
| | LMA.Eff.T-LHand | Sudden:**55%**, Sustained:45% |
| | LMA.Sh.V | Sinking:15%,Still:**75%**,Rising:10% |
| | LMA.R.T/A H-O1 | Toward:21%, Away:**79%** |
| | LMA.R.T/A H-O2 | Toward:**68%**, Away:32% |
| | LMA.R.C H-O1 | Connected:8%, Disconnected:**92%** |
| | LMA.R.C H-O2 | Connected:12%, Disconnected:**88%** |
| | Movement | Walk:**45%**, Stand:12%, Run:22%, Sit:8%,Rise:7%,Fall:6% |
| | Action and Interaction | Reaching:**48%**, Spreading:6%, Sit-on-chair:14%, Stand-up:10%, Handsh:4%,Pushing:6%,other:12% |
| 240 | | |
| | LMA.Eff.T-Head | Sudden:35%, Sustained:**65%** |
| | LMA.Eff.T-LFoot | Sudden:29%, Sustained:**71%** |
| | LMA.Eff.T-RFoot | Sudden:25%, Sustained:**75%** |
| | LMA.Eff.T-RHand | Sudden:**54%**, Sustained:46% |
| | LMA.Eff.T-LHand | Sudden:25%, Sustained:**75%** |
| | LMA.Sh.V | Sinking:25%,Still:**60%**,Rising:15% |
| | LMA.R.T/A H-O1 | Toward:45%, Away:**55%** |
| | LMA.R.T/A H-O2 | Toward:**55%**, Away:45% |
| | LMA.R.C H-O1 | Connected:11%, Disconnected:**89%** |
| | LMA.R.C H-O2 | Connected:**87%**, Disconnected:13% |
| | Movement | Walk:16%, Stand:**40%**, Run:10%, Sit:8%,Rise:16%,Fall:10% |
| | Action and Interaction | Reaching:8%, Spreading:10%, Sit-on-chair:10%, Stand-up:6%, Handsh:**37%**,Pushing:17%,other:12% |

Table IV

A TABLE WITH DIFFERENT LEVELS OF BODY-MOTION BASED INFORMATION FROM A PERSON WHO WALKS TO REACH OTHER PERSON AND DO HANDSHAKING (IN A COUPLE OF SEQUENCES) (SEE FIGURE 4-A)). FRAME NUMBER SHOWS THE LAST FRAME NUMBER OF THE WINDOW SLIDE (THE FIRST AND SECOND STEP CONTAINS THE FRAMES [60-180] AND [120-240], RESPECTIVELY). TABLE III EXPLAINS THE ACRONYMS.

| Frame | Level | States with their probability |
|---|---|---|
| 180 | | |
| | LMA.Eff.T-Head | Sudden:40%, Sustained:**60%** |
| | LMA.Eff.T-LFoot | Sudden:14%, Sustained:**86%** |
| | LMA.Eff.T-RFoot | Sudden:21%, Sustained:**79%** |
| | LMA.Eff.T-RHand | Sudden:25%, Sustained:**75%** |
| | LMA.Eff.T-LHand | Sudden:22%, Sustained:**78%** |
| | LMA.Sh.V | Sinking:10%,Still:**80%**,Rising:10% |
| | LMA.R.T/A H-O1 | Toward:51%, Away:**49%** |
| | LMA.R.T/A H-O2 | Toward:**78%**, Away:22% |
| | LMA.R.C H-O1 | Connected:8%, Disconnected:**92%** |
| | LMA.R.Ct H-O2 | Connected:18%, Disconnected:**82%** |
| | Movement | Walk:10%, Stand:**59%**, Run:8%, Sit:10%,Rise:7%,Fall:6% |
| | Action and Interaction | Reaching:**48%**, Spreading:6%, Sit-on-chair:14%, Stand-up:10%, Handsh:4%,Pushing:6%,other:12% |
| 240 | | |
| | LMA.Eff.T-Head | Sudden:**75%**, Sustained:25% |
| | LMA.Eff.T-LFoot | Sudden:**59%**, Sustained:41% |
| | LMA.Eff.T-RFoot | Sudden:**55%**, Sustained:45% |
| | LMA.Eff.T-RHand | Sudden:**84%**, Sustained:16% |
| | LMA.Eff.T-LHand | Sudden:**76%**, Sustained:24% |
| | LMA.Sh.V | Sinking:**75%**,Still:20%,Rising:5% |
| | LMA.R.T/A H-O1 | Toward:42%, Away:**58%** |
| | LMA.R.T/A H-O2 | Toward:**55%**, Away:45% |
| | LMA.R.C H-O1 | Connected:**87%**, Disconnected:13% |
| | Movement | Walk:10%, Stand:10%, Run:16%, Sit:18%,Rise:5%,Fall:**41%** |
| | Action and Interaction | Reaching:18%, Spreading:5%, Sit-on-chair:8%, Stand-up:5%, Handsh:17%,Pushing:**39%**,other:8% |

Table V

A TABLE WITH DIFFERENT LEVELS OF BODY-MOTION BASED INFORMATION FROM A PERSON WHO RUNS TO REACH OTHER PERSON AND PUSH HIM (IN A COUPLE OF SEQUENCES) (SEE FIGURE 4-B)). FRAME NUMBER SHOWS THE LAST FRAME NUMBER OF THE WINDOW SLIDE (THE FIRST AND SECOND STEP CONTAINS THE FRAMES [60-180] AND [120-240], RESPECTIVELY ). TABLE III EXPLAINS THE ACRONYMS.

LMA components, such as *Effort*.time of body parts and *Shape*.vertical, since their states are bipolar, learning process for one of the states was enough. In the human movement and interaction levels, half of the data, which were collected for each human movements and interactions, are used in learning process, and others for classification.

Table IV and Table V present the classification results of two window-slide sequences. As can be seen, different levels of information were presented semantically and probabilistically.

The overall result can be seen on the Table VI. The results prove that the context-based knowledge improves the accuracy of the model (from 92.22% to 96.80%) by reducing the false detections which was presented on the recent Khoshhal and Dias work [16]. As can be seen, between human-chair interactions and human-human ones there is no any false detection, however still there are some false detections between those similar context-based activities and especially between most of the classes with the *Other*'s class. Most of those false detections happen in the boundary of between two classes, because of using sliding window-based segmentation approach. When a slide window occur in the boundary, the new class of activity will be consider in the ground truth, though the window slide had more signal belong to the previous one.

## V. CONCLUSION AND FUTURE WORKS

In this study, a semantic-based hierarchical framework is proposed to deal with different levels of human movement analysis, based on the context-based information. The existent knowledge about the possible actions with respect to any object is a key of the system to improve the classification results. A well-known human movement descriptor , Laban movement analysis, was used to provide a standard description on body motion analysis level, which we call it LMA level. There are three components in the LMA level, *Effort* (deal with temporal domain), *Shape* (deal with spatial domain) and *Relationship* (deal with the context-based information), which are used in this study. Bayesian-based approach was used to model the multi-layer framework. In the output of the model a probabilistic-based descriptor for human activities was presented in different levels. Based on the result, it is proved that the context of a scene, where the human interaction is happening, can highly avoid the false detections, however still there are some false detections which happen in the boundary of two classes of activities.

We intend to apply the mentioned model on dataset from a smart-room which has a network camera, to detect people and objects [2], and using an adaptive based human movement segmentation to solve the mentioned drawback of this study.

| Acronym | Description |
|---|---|
| LMA.Eff.T-bp | LMA.Effort.Time component belongs to *bp*'s body part |
| LMA.Sh.V | LMA.Shape component in Vertical plane |
| LMA.R.T/A H-O | Toward/Away property of LMA.Relationship component between the person and object $O$. |
| LMA.R.C H-O | Contact property of LMA.Relationship component between the person and object $O$. |
| Handsh | Hand-shaking |

Table III
VARIABLES ACRONYMS

| | Reaching | Spreading | Sitting | Standing up | Hand shaking | Pushing | Other |
|---|---|---|---|---|---|---|---|
| Reaching | **97,78%** | 1,22% | 0,00% | 0,00% | 0,00% | 0,00% | 2,22% |
| Spreading | 0,00% | **95,74%** | 0,00% | 0,00% | 0,00% | 0,00% | 4,26% |
| Sitting | 0,00% | 0,00% | **96,67%** | 0,00% | 0,00% | 0,00% | 3,33% |
| Standing up | 0,00% | 0,00% | 2,70% | **97,30%** | 0,00% | 0,00% | 0,00% |
| Hand shaking | 0,00% | 0,00% | 0,00% | 0,00% | **97,83%** | 2,17% | 0,00% |
| Pushing | 0,00% | 0,00% | 0,00% | 0,00% | 2,27% | **97,73%** | 0,00% |
| Other | 1,09% | 2,17% | 1,09% | 1,09% | 0,00% | 0,00% | **94,57%** |

Table VI
CLASSIFICATION RESULTS

## VI. 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), and Jorge Dias is supported by the TICE Healthy Project (n.º 13842).

## REFERENCES

[1] J.K. Aggarwal and M.S. Ryoo. Human activity analysis: A review. *ACM Comput. Surv.*, 43(3):16:1–16:43, April 2011.

[2] Hadi Aliakbarpour and Jorge Dias. 3d reconstruction based on multiple virtual planes by using fusion-based camera network. *Journal of Computer Vision (IET)*, 2012.

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

[4] Min Chen, Sergio Gonzalez, Athanasios Vasilakos, Huasong Cao, and VictorC.M. Leung. Body area networks: A survey. *Mobile Networks and Applications*, 16(2):171–193, 2011.

[5] Diane Chi, Monica Costa, Liwei Zhao, and Norman Badler. The emote model for effort and shape. In *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 173 – 182, New York, NY, USA, 2000. Proceedings of the 27th annual conference on Computer graphics and interactive techniques, ACM Press/Addison-Wesley Publishing Co.

[6] Vincent Delaitre, David F. Fouhey, Ivan Laptev, Josef Sivic, Abhinav Gupta, and Alexei A. Efros. Scene semantics from long-term observation of people. In *Proceedings of the 12th European conference on Computer Vision - Volume Part VI*, ECCV'12, pages 284–298, Berlin, Heidelberg, 2012. Springer-Verlag.

[7] Magy Seif El-Nasr, Athanasios V. Vasilakos, C. Rao, and Joseph A. Zupko. Dynamic intelligent lighting for directing visual attention in interactive 3-d scenes. *IEEE Trans. Comput. Intellig. and AI in Games*, 1(2):145–153, 2009.

[8] 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.

[9] Abhinav Gupta, Aniruddha Kembhavi, and Larry S. Davis. Observing human-object interactions: Using spatial and functional compatibility for recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(10):1775–1789, October 2009.

[10] Ann Hutchinson. *Labanotation*. Oxford Univ. Press, 1974.

[11] K. Khoshhal, P. Menezes, and J. Dias. Descriptive human action and interaction analysis in a hierarchical framework. In *IROSW12 - International Workshop on Human-Agent Interaction (iHAI 2012)*, Vilamoura, Portugal, 2012.

[12] 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.

[13] 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)*, Delft, The Netherlands, April 2011. TU Delft; EWI; MM; PRB.

[14] Andrea Mannini and Angelo Maria Sabatini. Machine learning methods for classifying human physical activity from on-body accelerometers. *Sensors*, 10:1154–1175, 2010.

[15] Joerg Rett. *ROBOT-HUMAN Interface using LABAN Movement Analysis Inside a Bayesian framework*. PhD thesis, University of Coimbra, 2008.

[16] Kamrad Khoshhal Roudposhti and Jorge Dias. Probabilistic human interaction understanding - exploring relationship between human body motion and the environmental context. *Pattern Recognition Letters on Special Issue of Scene Understandings and Behaviours Analysis*, 2012.

[17] Kamrad Khoshhal Roudposhti, Luis Santos, Hadi Aliakbarpour, and Jorge Dias. Parameterizing interpersonal behaviour with laban movement analysis - a bayesian approach. In *CVPRW'12*, Providence, Rhode Island, USA, June 2012. IEEE.

[18] Liwei Zhao. *Synthesis and Acquisition of Laban Movement Analysis Qualitative Parameters for Communicative Gestures*. PhD thesis, University of Pennsylvania, 2001.