Gesture Recognition for Humanoid Robot Teleoperation

Insaf Ajili, Malik Mallem, Jean-Yves Didier

Gesture Recognition for Humanoid Robot Teleoperation

Ajili Insaf[1] Mallem Malik[2] and Didier Jean-Yves[3]

Abstract— **Interactive robotics is a vast and expanding research field. Interactions must be sufficiently natural, with robots having socially acceptable behavior by humans, adaptable to user expectations. Thus allowing easy integration in our daily lives in various fields (science, industry, health ...). Natural interaction during human-robot collaborative action needs suitable interaction techniques. In our paper we develop an online gesture recognition system for natural and intuitive communication between Human and NAO robot. However recognizing meaningful gesture patterns from whole-body gestures is a complex task. That is why we used the Laban Movement Analysis technique to describe high level gestures for NAO teleoperation. The major contributions of the present work is: (1) an efficient preprocessing step based on view invariant human motion, (2) a robust descriptor vector based on Laban Movement Analysis technique to generate compact and informative representations of Human movement, and (3) an online gesture recognition with Hidden Markov Model method was applied to teleoperate NAO based on our proper data base dedicated to the teleoperation of NAO. Our approach was evaluated with two challenging datasets, Microsoft Research Cambridge-12 (MSRC-12) and UTKinect-Action. Experimental results show that our approach outperforms the state-of-the-art methods.**

I. INTRODUCTION

Human gesture recognition in video sequences has become an emerging field of computer vision due to its importance in many practical applications such as human-computer interaction, multimedia event detection, video surveillance, robot learning and control, biometric security and health-care. The goal of human gesture recognition is to automatically analyze ongoing action from an unknown video. Human gesture recognition requires computer vision, machine learning techniques and especially the design of good feature representations. Although voluminous literature on the analysis and interpretation of human motion have emerged in recent years, the existing human action recognition features are still defective due to the high complexity of the human motion variability. More explicitly, for human robot interaction, robots should be able to recognize human gestures and identify different tasks defined by him. In this context, our paper consists in proposing an online recognition system for NAO teleoperation. Figure (1) illustrates the global idea of our work. For a successfull recognition system we need to faithfully describe Human gestures, which requires a robust motion descriptor. The major advantage in this step is to choose pertinent features while keeping our system independent of many factors that can happen in our working environment such as some geometric transformations (translation, scaling and view invariant) between different users. Our descriptor vector was derived from Laban Movement Analysis, a method for describing and interpreting all varieties of human movements. Three main components (Body, Space and Shape) have been defined. Next, extracted features are feeded into a Hidden Markov Model, to recognize command gestures sent to NAO. Finally our proposed method is evaluated with two public datasets and experimental results showed that our approach performs a high recognition rate.

The rest of the paper is organized as follows: in section II, we provide a brief review of the existing literature. In section III we define our proposed system as well as its various stages. In section IV we present our experimental evaluation, we describe an online user gesture recognition for robot teleoperation and conclude with some perspectives.

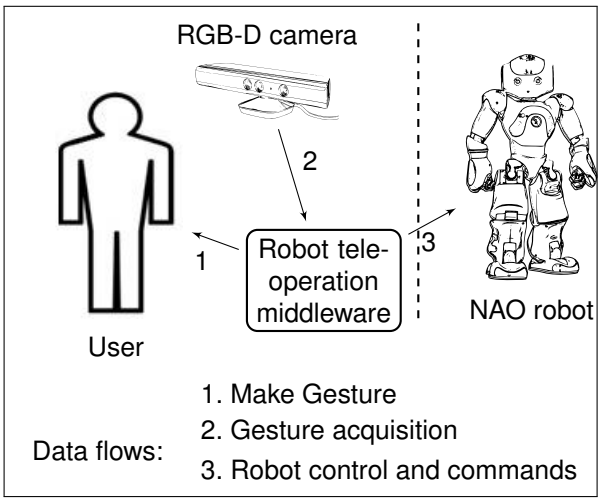

Fig. 1. Robot teleoperation setup

II. STATE OF ART

Teleoperation of humanoid robots is highly suitable for a number of applications especially complex tasks where human presence is impossible or dangerous, thus such a proposed system would be desirable. This kind of research touches many different areas, among them interaction modalities. Teleoperation can be performed with haptic controllers [19], [4], or by imitation learning with mapping the human motion to a robot [22], [8], [15], [1]. Other ways exist to make human robot interaction more natural, we talk about gesture recognition approach. Human gesture recognition has been extensively studied in the literature [7], [29], [14], [26], [24], [13], [12]. Previous researches have integrated

[1]Ajili Insaf [2]Mallem Malik and [3]Didier Jean-Yves are members in IBISC Lab (Insaf.Ajili, Malik.Mallem, Jean-Yves.Didier)@ufrst.univ-evry.fr

this approach as a human-robot interface for robot teleoperating. In [5] static and dynamic gesture recognition methods were applied for human multirobot interaction. Two features were chosen for dynamic gesture description, the euclidean distance between the hand and the neck joints and the angle in the elbow joint. Three other features were chosen for the static gesture presentation, such as the pointing gesture, the angle in the elbow joint, the distance between the hip and the hand joints, and the position of the hand joint. Weighted Dynamic Time Warping method has been applied to the dynamic gesture recognition. In [10] they recognized ten gestures chosen from the army visual signals to control a mobile robot. Gesture recognition was performed by neural network (NN) classifiers. They extracted two types of features (quaternions and angles) and compared between them to obtain a high recognition rate. In [11] they integrated the Euler angle to recognize left arm gestures to control the Pioneer robot with five gestures (come, go, wave, rise up and sit down). Their features are four joint angles (left elbow yaw and roll, left shoulder yaw and pitch).

Previous gesture recogniton researches show that most efforts have been devoted to the design of good feature gesture representations. Several types of descriptors have been chosen but most of them have focused on the quantitave aspect of gestures that does not suffisent to recognize high level gestures. We also need to know about the quality of gesture. An other research field was appear to treat this lack, the Laban Movement Analysis (LMA), one of the most famous theories of body expressions that was originally developped by a dancer, architect, choreographer and painter, Rudolf Laban [6]. This approach is used for representing and analyzing expressive movement and divided in four main components: Body, Space, Shape and Effort. Body quality describes structural and physical characteristics of body parts. The space category articulates the relationship between human body and three dimensional space. The shape category considers qualities describing the way the body changes shape during movement. This component is composed of three subcomponents (Shape Flow, Directional movement and carving). Finally effort component includes four subcategories: space, weight, time and flow. This category is always related to emotional analysis themes and body inner attitude. To our knowledge, only few papers proposed LMA approach and most of them for describing emotions or analyzing dance choreography. Some of them were interested only in two components, shape and effort, such as [28] in order to recognize players emotion for Exergames, and [23] for describing expressive qualities of hand and arm movements. In the other hand, others have designed a Laban descriptor vector composed of six features (Space, Time, Weight, Inclination, Height and Area) to achieve a correlation between Laban's feature and emotion expressed by movements, as [18], [20]. Our work consists in teleoperating NAO, which is supposed to recognize high level gestures and perform corresponding tasks. For this we should create an efficient recognition system that should be independent of many factors, such as user change view, user intention, gesture duration, etc. So if human user make the same gesture with two different speeds our system must recognize the same thing. For this we discarded effort component and we describe our gestures with three LMA qualities, body, space and shape.

III. PROPOSED PROCESS

Our proposed system consists in recognizing human gestures, it is constituted by three fundamental steps: Preprocessing data, Feature extraction and Recognition step.

A. Preprocessing data

In order to develop a consistent recognition system we first realized an efficient normalization step to make our system generic and functional in all circumstances (users's height, view invariant,...). We use Microsoft Kinect sensor, skeleton data contain 3D joint positions represented as $P_j(x_j, y_j, z_j)$, the position of the joint j in the camera coordinate system (X, Y, Z), X corresponds to the width, Y corresponds to the height and Z to the depth. Normalization stage consists of the following steps:

1. **Translation**, user can be found at any place whithin the coverage area of Kinect, and the coordinates of the same joint may assume different values. So we compensate the position of the skeleton by centering the coordinate space in one skeleton joint which is the midpoint (J_m) of the line connected the left hip joint (J_{lhi}) and the right hip joint (J_{rhi}) at initial frame. So we deplaced the midpoint (J_m) to the origin of the kinect coordinate system and we applied the same translation to all skeleton joints.
2. **Scaling**, to reduce the influence of different user's height, first the distance between J_m and the neck joint (J_n) of the subject was computed, then all skeleton 3D coordinates are normalized according to the calculated value.
3. **View invariant**, after translation and scaling we introduce a view-invariant description of human posture, we define a local skeleton coordinate system (X', Y', Z') with center is the midpoint of hips (J_m). We choose the hip axis whith direction from the right hip joint J_{rsh} to the left hip joint J_{lsh} as the X'-Axis of the skeleton coordinate system. The line passed between the midpoint of hips (J_m) and the spine joint (J_s) is treated as the Y' axis. And finally Z' is the orthogonal axis to X' and Y'. So the unit vectors of this new coordinate system are defined as follows:

$$i' = \frac{J_{lhi} - J_{rhi}}{\|J_{lhi} - J_{rhi}\|}; j' = \frac{J_s - J_m}{\|J_s - J_m\|}; k' = i' \wedge j' \quad (1)$$

Let $[J_i]_{(X,Y,Z)}$ be the 3D position of joint i presented in the camera coordinate system, we should compute the rotation matrix that can transform joints from camera coordinate system to skeleton coordinate system.

We have:

$$[J_i]_{(X,Y,Z)} = R_{(X,Y,Z)\leftarrow(X',Y',Z')}[J_i]_{(X',Y',Z')} \quad (2)$$
$$= [r_1\ r_2\ r_3][J_i]_{(X',Y',Z')} \quad (3)$$

where r_1, r_2 and r_3 are the rotation vectors around the $X-Axis$, $Y-Axis$ and $Z-Axis$ respectively.
By applying the equation (3), we have:
$r_1 = \frac{[J_{lhi}]_{(X,Y,Z)}}{a}$; $r_2 = \frac{[J_s]_{(X,Y,Z)}}{c}$ and $r_3 = r_1 \wedge r_2$ where a is the distance between J_m and J_{lhi}, and c is the distance between J_m and J_s. Since r_1, r_2 and r_3 are orthogonal unit vectors, hence $R^T = R^{-1}$. Then:

$$[J_i]_{(X',Y',Z')} = R^T_{(X,Y,Z)\leftarrow(X',Y',Z')}[J_i]_{X,Y,Z} \quad (4)$$

After the collection and calculation of initial transformations, we have to use the same transformations for each frame unless the kinect sensor is moved.

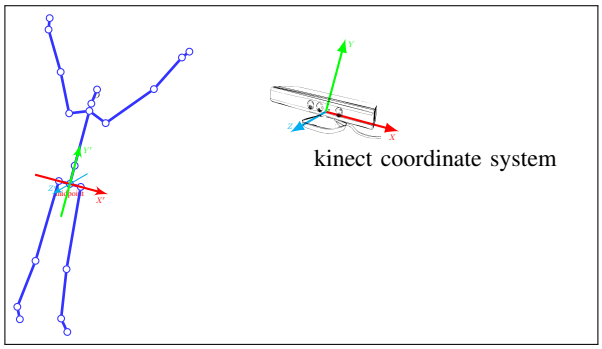

Fig. 2. Kinect coordinate system and skeleton coordinate system.

B. Feature extraction

Feature extraction is the most important step in a recognition system, good selection features improve classification performance. So, we have to find the most discriminating information for a faithful gesture representation. Our descriptor vector is derived from LMA technique [16]. Three components have participated in the description of our feature vector. Each component describes human motion by differents aspects.

Body component: defines structural and physical characteristics of the human body while moving. In this category we choose four features to describe how lower extremity joints (hands and foots) are organized in relationship to their upper extremity joints (shoulders and hips respectively). So we calculate angles between hand and shoulder joints as shown in Figure 3 and angle between foot and hip joints in the right and left part.

$$\theta_1^{l/r} = h^{l/r}\widehat{el^{l/r}sh^{l/r}}; \theta_2^{l/r} = f^{l/r}\widehat{kn^{l/r}hi^{r/l}} \quad (5)$$

Where $h^{l/r}$: Left/right hand joint, $el^{l/r}$: Left/right elbow joint, $sh^{l/r}$: Left/right shoulder joint, $kn^{l/r}$: Left/right knee joint, $f^{l/r}$: Left/right foot joint, and $h_i{}^{r/l}$ is the right/left hip joint.

We also compute the distance (d) between the hip center and the ground to know more about legs state if they are on the ground or jumping above it as shown in Figure 3.

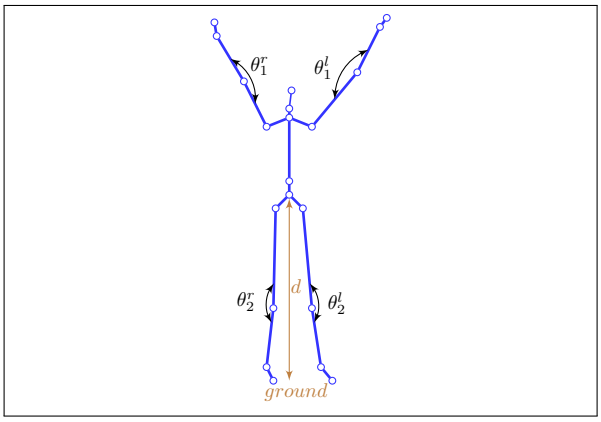

Fig. 3. Body features.

Space component: This category defines the interaction between the human body and the three dimensional space surrounding the body with directions spatial patterns and pathways. For the space category we define the torse direction by the normal vector $\vec{N}$ of triangle formed by left shoulder(sh_l), right shoulder (sh_r) and hip center (hi_c) joints) (Figure4). This vector gives the direction of the upper part of the skeleton during the movement:

$$\vec{N} = \frac{\overrightarrow{sh_l hi_c} \wedge \overrightarrow{sh_r hi_c}}{||\overrightarrow{sh_l hi_c} \wedge \overrightarrow{sh_r hi_c}||}$$

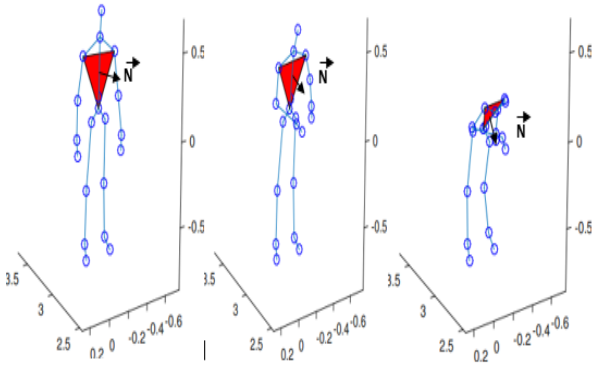

Fig. 4. Normal vector of triangle $\triangle$ (sh_l), (sh_r) and (sh_c) in "Take a bow" gesture from MSRC-12 Dataset [9].

Shape component: describes the changing forms that the body makes in space and charaterizes the body's interaction with itself as well as the space around it. Shape component is expressed by three factors of Shaping: Shape Flow, Directional Movement and Shaping Movement.

- Shape flow: analyses the changing forms that the body makes during movement. In this subcomponent we describe our skeleton body with two features. In the first one we construct the smallest convex envelope of the

skeleton human based on Quickhull algorithm [3] and we calculate its volume to characterize the deformation of the skeleton's bounding shape over time. Figure 5 illustrates the 3D skeleton convex hull and its shape deformation in "Crouch" gesture from the MSRC-12 database [9]. At initial frame the person is standing up, the volume of the enveloppe has the highest value, after user has crouched then the volume of the convex hull was decreased. By computing the volume of the convex hull we have more idea about body extension. But in this case we cannot know the source of this deformation, wether it comes from upper, lower or full body. For this reason we introduce another feature, the flattening index as a measure for the ratio between semi major and minor axis. The semi-major axis (a) is the line connecting two hands and the other semi-minor axis (b) is the segment between the shoulder center joint to the neck joint. Flattening index provides information about hands' extension in 3D space.

- Flattening index:

$$f = \frac{a-b}{a} \tag{6}$$

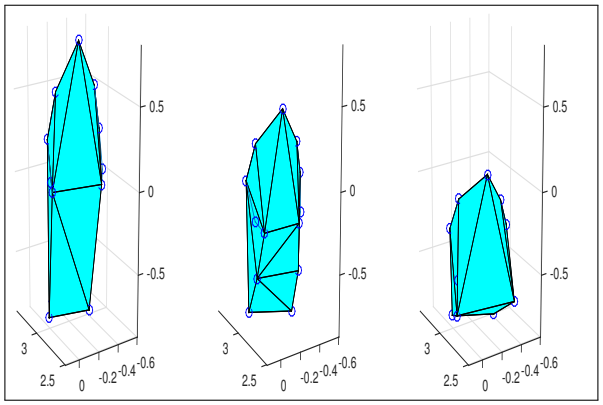

Fig. 5. Variation of convex hull feature in "Crouch" gesture from MSRC-12 Dataset [9].

- Directional movement: defines the pathway along which a movement is executed. These movements can be simple spoke-like or arc-like actions to reach a direction or object, such as touching an object or moving towards a specific location. For hands'pathway we define the local curvature of hands joints. So we consider the gradual angular change occuring between two frames, defined as follow (Figure 7):

$$\phi_{J_t} = \arccos \frac{\overrightarrow{J_{t-1}J_t} \cdot \overrightarrow{J_tJ_{t+1}}}{\left\|\overrightarrow{J_{t-1}J_t}\right\| \cdot \left\|\overrightarrow{J_tJ_{t+1}}\right\|} \tag{7}$$

where J_t is the position of hand joint at frame t and J_{t+1} is the position of hand joint at next frame $t+1$.

- Shaping Movement: depicts the way the shape of the human body evolves as the body interacts with the environment, that can be interpreted by projecting it to three planes: horizontal (Spreading/Enclosing movements), frontal (Rising, Sinking, movements) and sagittal plan

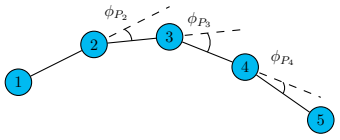

Fig. 6. Gradual angles between successives samples.

(Advancing/Retreating movements). We describe the displacement of head, upper and lower extremity joints in three planes (Figure 7). At each frame we compute projected distances between these joints relating to spine joint pose at initial frame. These projections provide the frontal, sagittal and horizontal view of head, arms and legs.

$$S = \sqrt{\sum_d (p_{ad} - p_{sd})} \tag{8}$$

where a represents each joint considered at each frame, s is the spine joint at initial frame and d belongs to one of the following sets $\{x, y\}, \{y, z\} and \{z, x\}$ for each considered projection.

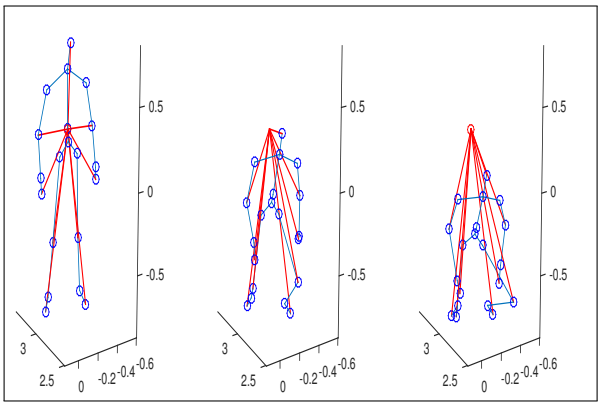

Fig. 7. Distances variation relating to spine joint, in "Crouch" gesture from MSRC-12 Dataset [9].

C. Recognition phase

In order to perform the feature classification of LMA features, we apply Discrete Hidden Markov Models [21] which accept as input discret values. We implement a C++ program to sample features sequences $f_k | k = 1, 2, \ldots, N$ into a fixed number of frames T. After sampling phase, feature vectors were quantized to a discrete symbols by applying K-means method [17].

IV. EXPERIMENTAL RESULTS

A. Offline action recognition

In the first experiment we would like to test the robustness of our descriptor vector with two challenging datasets MSRC-12, and UTkinect-action.

Algorithm 1: Sampling algorithm

function Sampling (f_k, T);
Input : $\{f_k | k = 1, 2, \ldots, N\}, T$
Output: $\{f_1, f_2, \ldots, f_T\}$
for $i \leftarrow 1$ **to** N **do**
  Calculate$D = \sqrt{\sum_{i=1}^{N}(f_i - f_{i+1})^2}$, distance between two successive feature vectors.
  if $d \leq \epsilon$ **then**
    $Delete f_{i+1}$
  end
end
Calculate the average distance: D'= $\frac{N'}{T}$ ▷ N' *is the number of feature vectors after removing noises.* ;
Get points after every d' feature vectors:
$f_{i+1} = f_i + d'$.

a) MSRC-12 Dataset [9]: is recorded using a Kinect sensor at $30Hz$. The dataset comprises 12 gesture classes, divided into two groups: iconic gestures and metaphoric gestures. The dataset has 594 sequences, collected from 30 subjects with different demographics. For every sequence, one subject performs an action several times and, for each frame, 3D locations of twenty body joints are estimated.

As a performance measure we apply 5×5 fold cross validation method [2]. It performs 5 replications of 5 fold cross validation. In each replication dataset is randomly divided into 5 equal sized sets, 4 subsets are used for training and one for testing. And the final classification results are averaged over the 5 random choices. The $F - score$ has been retained as performance measure by computing and combining precision and recall for each action.

$$F - score = 2 * \frac{precision * recall}{precision + recall}$$

$precision = \frac{T_p}{T_p + F_p}$; $recall = \frac{T_p}{T_p + F_n}$.

T_p is the number of true positives and F_n is the number of false negatives. We compare our framework with the baseline method [9], we make the same working conditions. We sample our data into 35 frames and we measure the intramodality generalization performance: training and testing using the same instruction modality for all gestures (iconic and metaphoric). The results in table I show that our approach outperforms the base model in all five modalities used for collecting the dataset. Also we compare our approach with [25] that proposed the same evaluation method. They have first performed the recognition separately on gestures of each given type (iconic, metaphoric). Then they have considered all gestures and they applied a 5-fold cross-validation repeated five times. As we can see in table Iour method outperforms [25] where the best F-scores obtained in the metaphoric gestures category was 81% and 79.77% in iconic category.

TABLE I
COMPARISON WITH BASELINE MODEL. MEAN AND AVERAGE F-SCORE OF ALL GESTURES CATEGORY FOR ALL THE FIVE MODALITIES.

Instructions	Text	Image	Video	I+T	V+T
Our approach	0.92 +/-0.05	0.96 +/-0.06	0.97 +/-0.02	0.96 +/-0.02	0.98 +/-0.01
Baseline [9]	0.621 +/-0.041	0.561 +/-0.034	0.673 +/-0.020	0.629 +/-0.051	0.709 +/-0.047

b) UTkinect [27]: is composed of 10 different subjects performing 10 different activities in varied views. A total number of 199 sequences are available. Each action is repeated twice by the actor. Sequences are captured using one Kinect in indoor settings and their length ranges from 5 to 120 frames. RGB, depth and skeleton joint locations are provided by Kinect for Windows SDK Beta Version, with a frame rate about 15 fps. This is a challenging dataset due to variations in the view point and high intra-class variations where each actor performs actions in different views. To compare our results with state of the art approaches, we follow experiment protocol proposed in [24] where they applied leave-one-out cross validation (LOOCV) [2] and achieved 95.25% accuracy. Our method outperform their recognition result with an average accuracy of 97%.

B. Online action recognition

After evaluating our system with complex datasets, now we want to test our recognition framework in real-time. We create a simple database dedicated to robot teleoperation composed of seven gestures (move forward, move backward, turn left, turn right, stand up, sit down, salute and stop). Ten subjects in our lab are selected to perform gestures. Each subject was asked to make gesture ten times. Our dataset has in total 700 sequences. For data acquisition, an OpenNI driver has provided a high level skeleton tracking module. This module required initial calibration to record 3D position of fifteen joints at 640×480 pixels. Our system is implemented under ROS, a robot operating system which consists in running a great number of executables to exchange data synchronously (via topics) or asynchronously (via services) as shown in Figure 8. We calculate our features from raw data by applying LMA, and then we make feature sequences into a fixed length by applying sampling algorithm with T=60 frames. After we cluster all feature vectors into discret values, we choose 12 as the number of symbols. Discrete HMMs with Left-Right Banded topology is applied for gesture modelisation and classification. To detect start and end gesture we verify if motion velocities v_x^j, v_y^j, et v_z^j of each joint j is close to zero for framelapse (we choose 25 frames).

As shown in Figures 9 and 10, Kinect camera captures both color images and depth. OpenNI tracking module publishes coordinates joints to gesture recognition modules (preprocessing, extraction, discretization, recognition, teleoperation,...). All defined gestures are recognized and converted to commands for NAO robot, such as "sit down"

gesture in Figure 9 and ”salut” gesture in Figure 10.

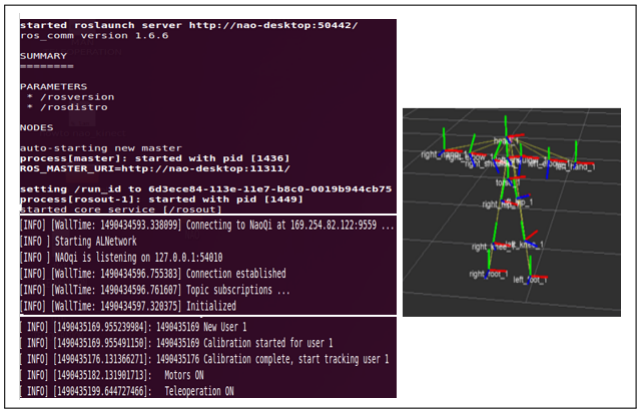

Fig. 8. Working space under ROS and skeleton visualization via a 3D visualization tool for ROS (RVIZ).

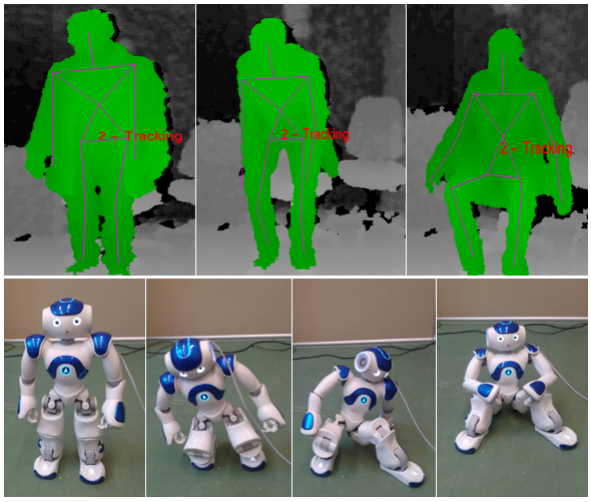

Fig. 9. Sit down gesture results on several sequences of frames applied to NAO.

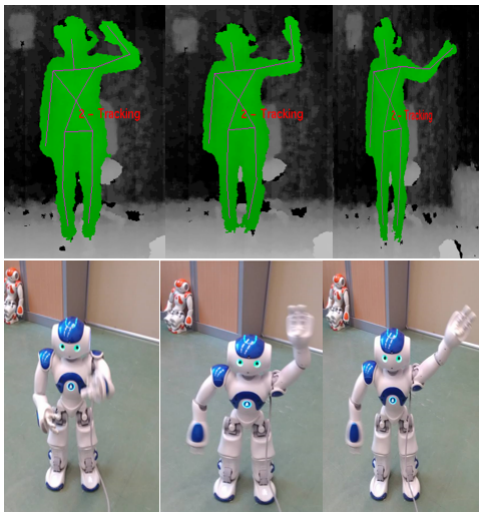

Fig. 10. Salute gesture results on several sequences of frames applied to NAO.

V. CONCLUSIONS

In this paper, we proposed a human gesture recognition system based on Laban Movement Analysis method. Three components derived from LMA are extracted, Body, Shape and Space in order to describe our feature vector. To realize automatic gesture recognition, preprocessing data, extraction features, sampling and discretization method were developed. Then HMMs with Left-Right Banded topology were used to model and classify gestures.

Experimental results on two challenging datasets MSRC-12 and UTkinect-action confirm the right choice of our descriptor vector. As perspectives, we plan to improve our database with gestures more complex, and integrate Simultaneous Localization and Mapping (SLAM) method in order to teleoperate NAO and explore environment.

REFERENCES

[1] I. Almetwally and M. Mallem. Real-time tele-operation and tele-walking of humanoid robot nao using kinect depth camera. In *2013 10th IEEE INTERNATIONAL CONFERENCE ON NETWORKING, SENSING AND CONTROL (ICNSC)*, pages 463–466, April 2013.

[2] S. Arlot and A. Celisse. A survey of cross-validation procedures for model selection, 2009.

[3] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa. The quickhull algorithm for convex hulls. *ACM Trans. Math. Softw.*, 22(4):469–483, Dec. 1996.

[4] A. Brygo, I. Sarakoglou, N. Garcia-Hernandez, and N. Tsagarakis. *Humanoid Robot Teleoperation with Vibrotactile Based Balancing Feedback*, pages 266–275. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014.

[5] G. Canal, C. Angulo, and S. Escalera. Gesture based human multi-robot interaction. In *2015 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, July 2015.

[6] A. L. de Souza. *Laban Movement Analysis—Scaffolding Human Movement to Multiply Possibilities and Choices*, pages 283–297. Springer International Publishing, Cham, 2016.

[7] K. M. Deo, A. Y. Kulkarni, and T. S. Girolkar. Hand gesture recognition using colour based segmentation and hidden markov model. *International Journal of Computer Applications Technology and Research*, 4(5):386–389, 2015.

[8] S. Filiatrault and A. M. Cretu. Human arm motion imitation by a humanoid robot. In *2014 IEEE International Symposium on Robotic and Sensors Environments (ROSE) Proceedings*, pages 31–36, Oct 2014.

[9] S. Fothergill, H. Mentis, P. Kohli, and S. Nowozin. Instructing people for training gestural interactive systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’12, pages 1737–1746, New York, NY, USA, 2012. ACM.

[10] C. G. T. D. Grazia Cicirelli, Carmela Attolico. A kinect-based gesture recognition approach for a natural human robot interface. *International Journal of Advanced Robotic Systems*, 12, March 19, 2015.

[11] Y. Gu, H. Do, Y. Ou, and W. Sheng. Human gesture recognition through a kinect sensor. In *Robotics and Biomimetics (ROBIO), 2012 IEEE International Conference on*, pages 1379–1384, Dec 2012.

[12] M. E. Hussein, M. Torki, M. A. Gowayyed, and M. El-Saban. Human action recognition using a temporal hierarchy of covariance descriptors on 3d joint locations. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, IJCAI ’13, pages 2466–2472. AAAI Press, 2013.

[13] X. Jiang, F. Zhong, Q. Peng, and X. Qin. Online robust action recognition based on a hierarchical model. *Vis. Comput.*, 30(9):1021–1033, Sept. 2014.

[14] I. N. Junejo, K. N. Junejo, and Z. A. Aghbari. Silhouette-based human action recognition using sax-shapes. *The Visual Computer*, 30(3):259–269, 2014.

[15] J. Koenemann, F. Burget, and M. Bennewitz. Real-time imitation of human whole-body motions by humanoids. In *2014 IEEE Interna-*

tional Conference on Robotics and Automation (ICRA), pages 2806–2812, May 2014.

[16] R. Laban. *La Maîtrise du mouvement*. Actes Sud Beaux Arts, 1994.

[17] J. Macqueen. Some methods for classification and analysis of multivariate observations. In *In 5-th Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297, 1967.

[18] M. Masuda, S. Kato, and H. Itoh. *A Laban-Based Approach to Emotional Motion Rendering for Human-Robot Interaction*, pages 372–380. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[19] N. C. Mitsou, S. V. Velanas, and C. S. Tzafestas. Visuo-haptic interface for teleoperation of mobile robot exploration tasks. In *ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication*, pages 157–163, Sept 2006.

[20] K. Nishimura, N. Kubota, and J. Woo. Design support system for emotional expression of robot partners using interactive evolutionary computation. In *2012 IEEE International Conference on Fuzzy Systems*, pages 1–7, June 2012.

[21] L. R. Rabiner and B. H. Juang. An introduction to hidden Markov models. *IEEE ASSP Magazine*, pages 4–15, January 1986.

[22] J. Rosado, F. Silva, and V. Santos. A kinect-based motion capture system for robotic gesture imitation. In M. A. Armada, A. Sanfeliu, and M. Ferre, editors, *ROBOT (1)*, volume 252 of *Advances in Intelligent Systems and Computing*, pages 585–595. Springer, 2013.

[23] A. A. Samadani, S. Burton, R. Gorbet, and D. Kulic. Laban effort and shape analysis of affective hand and arm movements. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, pages 343–348, Sept 2013.

[24] R. Slama, H. Wannous, and M. Daoudi. Grassmannian representation of motion depth for 3d human gesture and action recognition. In *2014 22nd International Conference on Pattern Recognition*, pages 3499–3504, Aug 2014.

[25] Y. Song, L. P. Morency, and R. Davis. Distribution-sensitive learning for imbalanced datasets. In *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, pages 1–6, April 2013.

[26] Y. Song, J. Tang, F. Liu, and S. Yan. Body surface context: A new robust feature for action recognition from depth videos. *IEEE Transactions on Circuits and Systems for Video Technology*, 24(6):952–964, June 2014.

[27] L. Xia, C. C. Chen, and J. K. Aggarwal. View invariant human action recognition using histograms of 3d joints. In *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 20–27, June 2012.

[28] H. Zacharatos, C. Gatzoulis, Y. Chrysanthou, and A. Aristidou. Emotion recognition for exergames using laban movement analysis. In *Proceedings of Motion on Games*, MIG '13, pages 39:61–39:66, New York, NY, USA, 2013. ACM.

[29] M. Zanfir, M. Leordeanu, and C. Sminchisescu. The moving pose: An efficient 3d kinematics descriptor for low-latency action recognition and detection. In *2013 IEEE International Conference on Computer Vision*, pages 2752–2759, Dec 2013.