Applying a 3D Qualitative Trajectory Calculus to Human Action Recognition using Depth Cameras

Claudio Coppola,[1] Oscar Martinez Mozos,[1] Nicola Bellotto[1]

***Abstract*— The life span of ordinary people is increasing steadily and many developed countries are facing the big challenge of dealing with an ageing population at greater risk of impairments and cognitive disorders, which hinder their quality of life. Monitoring human activities of daily living (ADLs) is important in order to identify potential health problems and apply corrective strategies as soon as possible. Towards this long term goal, the research here presented is a first step to monitor ADLs using 3D sensors in an Ambient Assisted Living (AAL) environment. In particular, the work here presented adopts a new 3D Qualitative Trajectory Calculus (QTC_{3D}) to represent human actions that belong to such activities, designing and implementing a set of computational tools (i.e. Hidden Markov Models) to learn and classify them from standard datasets. Preliminary results show the good performance of our system and its potential application to a large number of scenarios, including mobile robots for AAL.**

I. INTRODUCTION

The life span of ordinary people is increasing steadily and many developed countries are facing the big challenge of dealing with an ageing population at greater risk of impairments and cognitive disorders, which hinder their quality of life. According to the United Nations [1], 2 billion persons are expected to be over 60 by 2050. This increment in the elderly population is accompanied by future shortages of available health workers and doctors as stated by the Organization for Economic Co-operation and Development [2]. Thus, monitoring human activities of daily living (ADLs) is important in order to identify potential health problems and apply corrective strategies as soon as possible.

Towards this long term goal, the research here presented is a first step to monitor ADLs from depth data in an Ambient Assisted Living (AAL) environment. In particular, we consider human actions captured by an RGB-D camera, where a skeleton representation of the body is extracted from depth images and its joints tracked over time, as shown in Fig. 1. The spatio-temporal interaction between different joints is captured by applying a new 3D Qualitative Trajectory Calculus (QTC_{3D}) [3], which allows to represent and classify different human actions in a domestic environment. Using QTC_{3D} relations of body joints, we then learn Hidden Markov Models (HMMs) to represent each actions as a sequence of joint interactions.

Spatial relations between body joints have been used as a feature for action recognition in the past [4], however only relative distance between joints was used as a spatial feature.

[1]Lincoln Centre for Autonomous Systems (L-CAS), School of Computer Science, University of Lincoln, LN6 7TS, Lincoln, United Kingdom {ccoppola, omozos, nbellotto}@lincoln.ac.uk

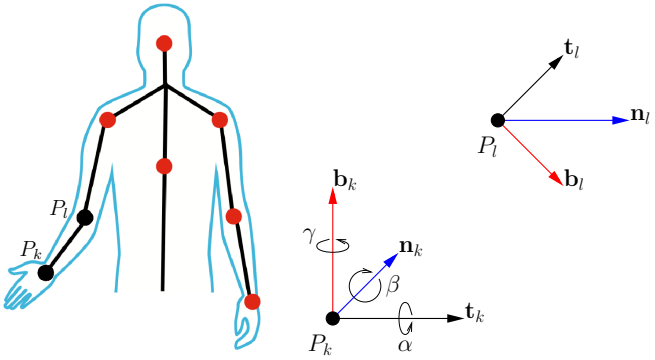


Fig. 1: Human skeleton (left), and Discrete Frenet Frames of two body joints P_k and P_l (right).

In contrast, in this work we extend these relations between joints using QTC_{3D}, which provides a richer representation of relative spatio-temporal relations such as "joint A approaches joint B" or "joint A is stable with respect to joint B." Finally, the different human actions are encoded by sequences of QTC_{3D} states that are represented using HMMs. The final trained HMMs are then used to classify new actions carried out by persons. Two main contributions are therefore provided by this paper:

- The first application of QTC_{3D} to the problem of human action recognition from depth data;
- The development of computational tools for the practical implementation and evaluation of such application.

In this work, we have applied our approach to classify actions from ADLs that are selected taking into account the ability groups, such as feeding and transferring, defined in widely used assessment indexes for carers and nursing homes. Preliminary results show the good performance of our system and its potential application to a large number of scenarios, including mobile robots for AAL.

The reminder of the paper is as follows: Sec. II reviews some relevant qualitative and numerical solutions to human action recognition, including depth-only and skeleton-based approaches; Sec. III presents the original QTC formalism and its extension to QTC_{3D}; Sec. IV explains the actual implementation of the HMMs used for action recognition; Sec. V describes the experiments and illustrates the results for several ADLs; finally, Sec. VI discusses advantages and current limitations of the proposed methods, as well as directions for future research work.

II. RELATED WORK

Action recognition is a well established research area. Different approaches, mostly vision-based, have been pro-

posed in the past to solve this problem [5]. With the recent introduction of depth sensors and related algorithms for body parts tracking [6], new powerful solutions for 3D action recognition have being devised. In this context, a significant number of approaches based on Qualitative Spatial Representations and Reasoning (QSR) [7] have become relevant.

In [8], for example, an activity graph-based approach is used alongside several possible qualitative representations for describing video activities. The work in [9] proposed a model for activity monitoring based on QSR, generative models and HMMs. The work in [10] shows an application of the Laban Movement Analysis as a qualitative categorization of motion behaviours to label human emotion from movements, which are observed with a stereo vision system. Recently, authors in [11] presented an application of a 2D Qualitative Trajectory Calculus (QTC)to model human-robot spatial interactions, where human actions consisted of walking patterns.

The work in [12] proposes a qualitative representation of human motion based on "Posebits" features that describe the variation of distance, articulation angles or relative position of body parts. In [13], instead, actions are represented using body motion attributes, which are partly pre-defined by an expert and partly obtained by real data on 2D images, and modelled as latent variables to select the most discriminative set. The authors in [14] adopt 2D trajectory descriptors to represent body movements as sequences of joints displacements, including relational features, to describe joints distances and orientations in each frame. Since they are based on 2D data only, the last two approaches are view-dependent.

A large number of numerical rather than qualitative approaches are based on depth data from RGB-D cameras. The work proposed in [15], for example, uses random occupancy patterns for recognising actions from depth data. In [16], the dynamics of the actions are modelled into an action graph, which is trained on a bag of 3D points. The authors in [17] propose a solution that generates 4D normals using depth, time, and spatial coordinates to create a histogram which encodes the distribution of the surface normal orientation. Another system is presented in [18] that generates motion maps by accumulating the differences between depth frames, and then used Histogram of Oriented Gradients (HOG) to describe those maps. A support vector machine (SVM) is used to classifying human actions.

Many recent solutions perform action recognition based on the body skeleton and joints that is possible to extract from depth data. In [19], for example, features are collected in random sub-volumes around the local area surrounding skeletal joints using the Iterative Signature Algorithm.

The work presented in [4] uses skeleton-based features are enriched with local occupancy patterns to improve the classification, and a final pool of informative "actionlets" are obtained through a mining process. In [20], eigen-joints are obtained from joint position differences.

In [21], instead, joints are grouped to obtaining an orientation invariant descriptor. The joints are modelled with a Gaussian Mixture Model and encoded as a Fisher vector, which is used for action classification with a one-vs-all SVM. Finally, the authors in [22] propose a view invariant descriptor based on histograms of 3D joint positions, voted with a Gaussian weight function. Features are extracted with Linear Discriminant Analysis to obtain a series of symbols, which are classified with a HMM.

Our approach, based on qualitative spatio-temporal representations of body joints, provides an alternative solution, powerful yet simple, to the aforementioned systems. It is based on a recent extension of QTC that, in previous versions, was limited to 2D motion only and therefore was unsuitable for 3D action recognition. Extending this representation to 3D allows us to be view-independent for future applications in mobile robotics.

III. 3D QUALITATIVE TRAJECTORY CALCULUS

The Qualitative Trajectory Calculus (QTC) is a mathematical formalism to represent qualitative information about moving objects, specifically considering relative spatio-temporal relations between couples of moving points [23]. There are different variants of this calculus, mostly dealing with points in 2D. For the purpose of this paper, however, we consider a recent extension of QTC that represents the relative motion between 3D points, including relations about distance and orientation, called QTC_{3D} [3]. We start by introducing a simple version of QTC in 2D, where the qualitative relations between two moving points P_k and P_l are expressed by the symbols $q_i \in \{-, +, 0\}$ as follows:

q_1) $-$: P_k is moving towards P_l;
0: P_k is stable with respect to P_l;
$+$: P_k is moving away from P_l;
q_2) same as q_1, but swapping P_k and P_l;
q_3) $-$: P_k is moving to the left side of $\overrightarrow{P_kP_l}$;
0: P_k is moving along $\overrightarrow{P_kP_l}$;
$+$: P_k is moving to the right side of $\overrightarrow{P_kP_l}$;
q_4) same as q_3, but swapping P_k and P_l.

A string of QTC symbols $\{q_1, q_2, q_3, q_4\}$ is a compact representation of the relative motion in 2D between P_k and P_l. For example, $\{-, -, 0, 0\}$ could be read as "P_k and P_l are moving straight towards each other". Obviously, the side relations cannot equally apply to the 3D case, therefore q_3 and q_4 have been replaced in QTC_{3D} by three new symbols that consider the relative roll, pitch and yaw of the frames $\mathbf{F}_k$ and $\mathbf{F}_l$ associated to P_k and P_l respectively. In particular, if $\mathbf{x}_k(t)$ is the positions of point P_k at (discrete) time t, then $\mathbf{F}_k$ is the Discrete Frenet Frame [24] identified by the following tangent $\mathbf{t}_k$, normal $\mathbf{n}_k$ and binormal $\mathbf{b}_k$ (see Fig. 1):

$$\begin{aligned} \mathbf{t}_k(t) &= \frac{\mathbf{x}_k(t+1) - \mathbf{x}_k(t)}{|\mathbf{x}_k(t+1) - \mathbf{x}_k(t)|} \\ \mathbf{b}_k(t) &= \frac{\mathbf{t}_k(t-1) \times \mathbf{t}_k(t)}{|\mathbf{t}_k(t-1) \times \mathbf{t}_k(t)|} \\ \mathbf{n}_k(t) &= \mathbf{b}_k(t) \times \mathbf{t}_k(t) \end{aligned} \tag{1}$$

Similarly, $\mathbf{t}_l$, $\mathbf{n}_l$ and $\mathbf{b}_l$ define a frame $\mathbf{F}_l$ for point P_l. We can compute then a rotation matrix $\mathbf{R}$ that aligns $\mathbf{F}_k$ with $\mathbf{F}_l$ as follows:

$$\mathbf{F}_l = \mathbf{R}\mathbf{F}_k \;\Rightarrow\; \mathbf{R} = \mathbf{F}_l\mathbf{F}_k^{-1} \tag{2}$$

If r_{ij} is the element in the i-th row and j-th column of $\mathbf{R}$, the roll α, the pitch β and the yaw γ necessary to align $\mathbf{F}_k$ with $\mathbf{F}_l$ can be calculated as follows [25]:

$$\begin{aligned} \alpha &= \tan^{-1}(r_{21}/r_{11}) \\ \beta &= \tan^{-1}(-r_{31}/\sqrt{r_{32}^2 + r_{33}^2}) \\ \gamma &= \tan^{-1}(r_{32}/r_{33}) \end{aligned} \tag{3}$$

where, by considering the quadrant of the inverse tangent's argument, all the three angles are within the interval $(-\pi, \pi]$.

The signs of α, β and γ lead to the three new qualitative symbols, q_5, q_6 and q_7 respectively, where the latter assume the value $-$, 0 or $+$ depending on whether the angle is less than, equal to or greater than zero. Note that, for practical reasons, a threshold ϵ_{angle} is set around zero so that very small angles are indeed considered qualitatively 0. A similar approach is tipically applied to the distance used to compute q_1 and q_2, for which a threshold ϵ_{dist} around zero is also used. A method for empirically tuning these thresholds is discussed in [26]. Finally, the string $s_{kl} = \{q_1, q_2, q_5, q_6, q_7\}$, replacing q_3 and q_4 with the newly create symbols, constitutes the QTC$_{3D}$ relation between points P_k and P_l.

IV. QTC$_{3D}$-BASED HIDDEN MARKOV MODELS

The objective of this work is to be able to classify different activities that people carry out during their daily life. For this purpose we use an ADLs dataset containing daily activities of a person recorded with an RGB-D camera. Each example in this dataset consists of a sequence of depth images captured while the person performs a specific action, e.g. *eating*. Example images for different actions are shown in Fig.2. Each image in the dataset includes the skeleton of the person performing the action, including the positions of the joints, also shown in the Fig. 4 These joints are used to analyse human movements. At every time instant, QTC$_{3D}$ strings are associated to each possible couple of joints to form the current body state; temporal state sequences are analysed then to recognise the action performed by the subject.

Formally, an action $\mathcal{A}$ carried out by a person, e.g. *eating*, is composed of a sequence of body positions $S(t)$ assumed by the person during the action, i.e. $\mathcal{A} = \{S(1), ..., S(T)\}$, where T is the number of depth images for the action in the dataset. In our qualitative representation, each body position S is defined by a set of QTC$_{3D}$ strings between all the possible couples of joints (P_k, P_l), that is $S = \{s_{kl} \mid k = 1, ..., N \wedge l = 1, ..., N \wedge k \neq l \wedge kl \neq lk\}$, where N is the total number of joints and each couple is represented only once, i.e. (P_k, P_l) and (P_l, P_k) are equivalent. As shown in (1), QTC$_{3D}$ uses temporal information to calculate the relative motion between joints. Therefore, to compute $S(t)$ we need the joints' position from the depth images at time $t-1$, t and $t+1$.

In our system, the classification is performed by several independent banks B_i of HMMs, each one corresponding to a particular action $\mathcal{A}_i$. The bank B_i contains one HMM for each possible couple of joints, which are $H = |S|$ in total. This number can be easily calculated as $H = \binom{N}{2} = \frac{N!}{2!(N-2)!}$, where N is the number of joints. This structure allows us to add new actions to the recognition system without need to retrain previous banks. The same applies to new joints, which can be easily added by incorporating additional HMMs trained separately with new QTC$_{3D}$ string sequences.

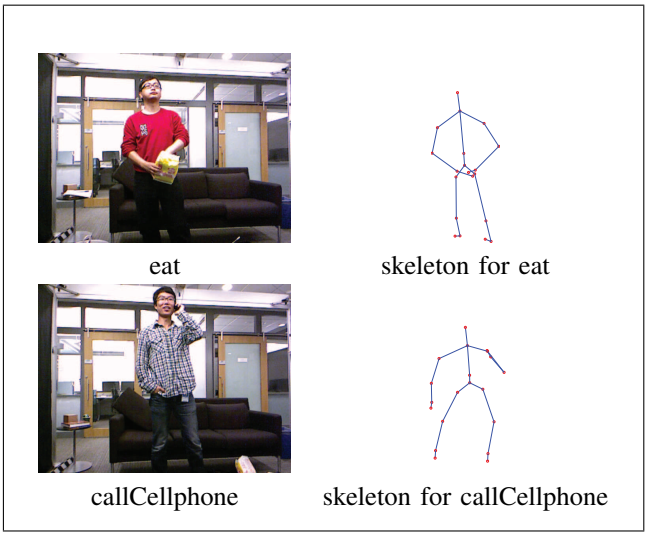


Fig. 2: Examples of actions from ADLs dataset [4] captured by an RGB-D camera, including skeleton and joints.

A. Clustering and Training

Following existing literature, the length of each HMM (i.e. number of hidden states) is fixed a priori. We make use of this property to initialize the transition and emission matrices of the HMMs based on the following clustering method applied to QTC$_{3D}$ training strings. Differently from other standard solutions (i.e. uniform matrix initialization), our approach is fully repeatable and, in our experience, does always lead to fast convergence of the HMM training algorithm.

Clustering is done only once before actual training of the HMMs to initialize the respective transition and emission matrices. For this purpose, we consider an action $\mathcal{A}_i$ and a particular couple of joints (P_k, P_l), for which there is a specific HMM in the respective bank B_i. We take then all the possible QTC$_{3D}$ strings $s_{kl}^i(t)$, with $t = 1, ..., T$, for all the training data sequences (i.e. depth videos) relatively to action $\mathcal{A}_i$. After removing duplicates, we group them in $C_1, ..., C_M$ clusters, where M is the number of hidden states of the HMM. To make this initialization process be driven by data only, we use agglomerative clustering [27]. The algorithm starts with one cluster for each sample; at every iteration, it merges the clusters that are close to each other, until it reaches a stopping condition, which in our case is the fixed number M of clusters (i.e. same as hidden states).

The distance used by the clustering algorithm is a standard one used in most QTC applications, given by the sum of the distances between corresponding symbols in two different QTC$_{3D}$ strings. So, if a and b are two QTC$_{3D}$ strings of length n, a_j and b_j their j^{th} symbols, and $d_q(a_j, b_j)$ the

distance between the two symbols defined as follows:

$$d_q(a_j, b_j) = \begin{cases} 0 & a_j = b_j \\ 1 & a_j = +, b_j = 0 \\ 1 & a_j = -, b_j = 0 \\ 2 & a_j = -, b_j = + \end{cases} \quad (4)$$

then the total distance between a and b is simply given by the following sum:

$$d_{QTC}(a, b) = \sum_{j=1}^{n} d_q(a_j, b_j) \quad (5)$$

The actual initialization of the transition matrix, having size $M \times M$, consists in counting all the transitions from elements of cluster C_m to elements of cluster C_n, and normalizing across all the transitions from C_m to any other cluster. Similarly, the initialization of the the emission matrix, having size $M \times Q$ (where $Q = 3^5$ is the total number of possible QTC_{3D} strings), is performed by counting all the training strings with identical values associated to a particular cluster C_m. Normalization in this case is done across all the possible strings associated to the same cluster.

The above initialization procedure is repeated for all the HMMs (i.e. all the couples of joints) of the current bank B_i relatively to action $\mathcal{A}_i$, and then again for all the other banks relatively to other actions. After that, the actual training of each HMM is done by refining the initial matrices with the Viterbi Algorithm and QTC_{3D} string sequences.

B. Classification

Some previous works have considered coupled or linked HMMs to compute an action probability for each single joint [28]. In our case though, a relatively simple and straightforward approach based on independence and product of probabilities has proved to be successful in practice.

Every bank B_i is made of H separate HMMs, one for each couple of joints (P_k, P_l). Given a new QTC_{3D} string sequence $s^i_{kl}(t)$, the respective HMM returns a probability of action $\mathcal{A}_i$ for the particular couple of joints. Assuming independence of the latter, a total probability for the bank can be computed as the product of the single HMMs probabilities (Fig. 3a). That is, if $p_h(\mathcal{A}_i)$ is the output probability of a single HMM, then $p_B(A_i) = \prod_{h=1}^{H} p_h(\mathcal{A}_i)$ is the total probability for B_i (normalized across all the actions / banks of HMMs)

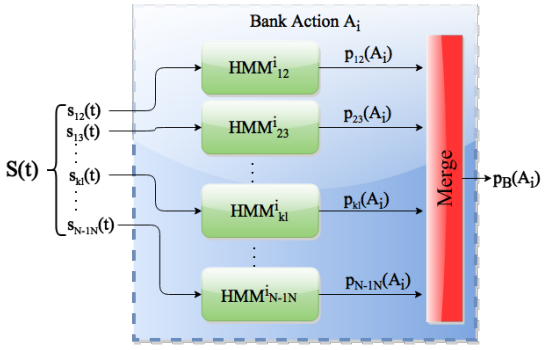
Eventually, the action $\mathcal{A}_j$ selected by the classification is the one that maximize p_b (Fig. 3b), that is:

$$j = \arg\max_i p_B(\mathcal{A}_i) \quad (6)$$

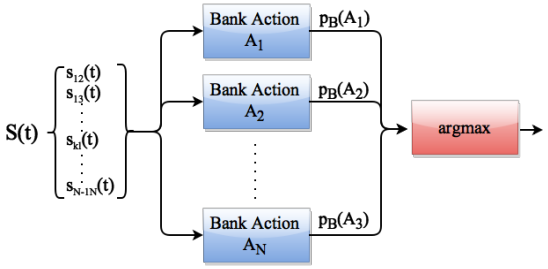
V. EXPERIMENTS

To test the validity of our approach we have designed a set of experiments in which we applied our QTC_{3D} representation together with HMMs to classify different sequences of actions corresponding to various ADLs.

In our experiments we used the MSR Daily Activity 3D dataset [4], which contains human actions captured using a



(a) Structure of the single Bank of Hidden Markov Models HMM^i_{kl}, each associated to action i and joints (k, l) to compute the probability of an action being occurred $p_B(A_i)$.



(b) Selection of the action A_j which has max probability $p_B(A_j)$.

Fig. 3: Architecture of the classification system: In input the sequences of QTC_{3D} strings $s^i_{kl}(t)$ and as output the action A_j which has max probability of being occurred.

Kinect camera. This dataset provides also the skeleton and joints of the persons performing the activities. The dataset contains 16 different activities carried out in a living room. Since the aim of this work is to monitor activities for daily living, we have selected a subset of actions that are directly related to ADL. We based our selection on the ability groups presented in the Bristol activities of daily living scale [29] and the Katz index of independence in activities of daily living [30]. In particular, our subset of actions includes *eating*, which is related to the feeding activity; *laying down on sofa*, *sitting down*, *sitting still*, *standing up*, and *walking*, which are related to mobility; the *calling cellphone* action, which is related to more general telephoning activity; and finally *using vacuum cleaner*, as part of housekeeping. Example frames for the previous activities are shown in Fig. 4.

Each one of the selected eight actions is carried out by 10 different people. Every person performs the action twice: once while sitting on the sofa, and once by standing still. In total our dataset contains 160 sequences. Each sequence represents a specific action carried out by one person in one of the two states (sitting or standing).

For the rest of this section we apply the following parameters in our approach. For the conversion from skeletal joints coordinates to QTC_{3D} strings we apply the thresholds $\epsilon_{dist} = 10^{-2}$ m and $\epsilon_{angle} = 3 \cdot 10^{-3}$ rad, which were empirically determined to obtain good results. The number of clusters used for the HMM creation was set to 5, since this valued provided the better classification results. In the Viterbi

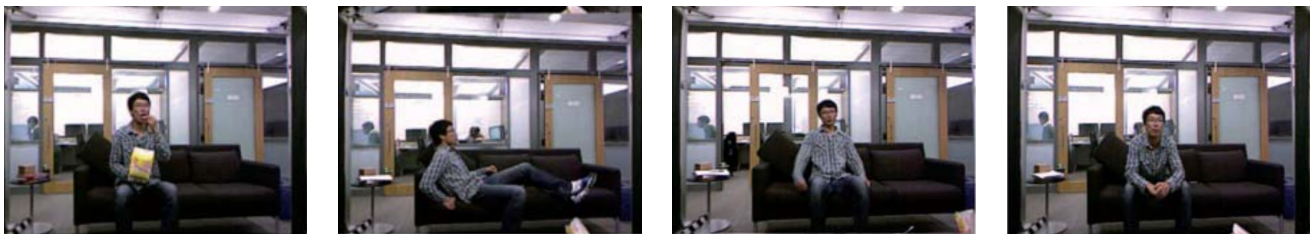


Fig. 4: Example frames of selected activities from the MSR Daily Activity 3D dataset [4].

algorithm used for training the HMMs we set a maximum of 100 iterations and a tolerance value for convergence of 10^{-4}, which are the standard values used in similar systems and always lead to convergence in our experiments. Furthermore, pseudo-emissions and pseudo-transitions are enabled so the minimum probability of occurring a transition/emission is 10^{-15}. This low probability allows us to overcome the problem of a lack of sufficient amounts of training data and unobserved transitions therein [11].

In the first experiment we applied a leave-one-out cross-validation approach to test our classifier. In this approach, we selected one sequence as test and used the rest of the 159 sequences to train our HMMs. We repeated this process 160 times, each time selecting a different sequence as test. The resulting confusion matrix is shown in Fig. 5. The average correct classification rate (using the average over the diagonal) for this experiment was 75.00% . As we can see in the confusion matrix in Fig. 5 most the actions are detected with high accuracy. However, our system tends to confuse the actions *eat* and *callCellphone*. We guess this confusion is due to the similar movement (hand approaches the head) that is carried out during both actions. A solution to this problem would consist on analysing the object that is held by the person as suggested in [4].

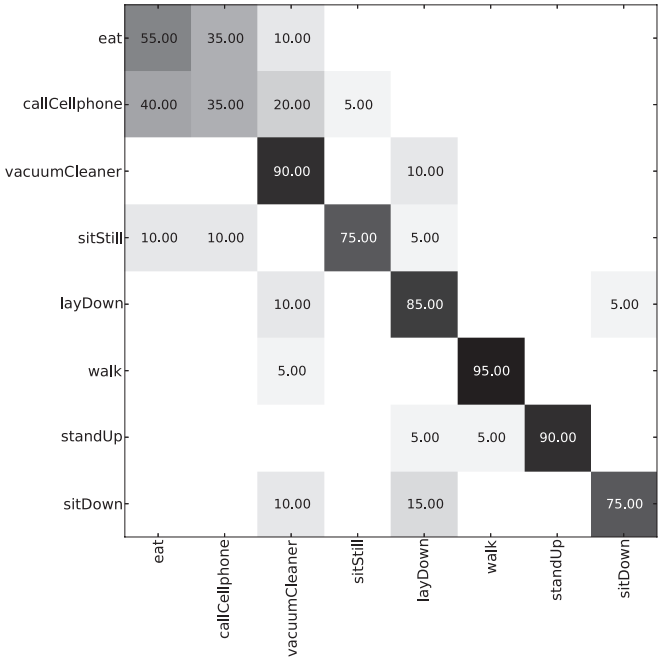


Fig. 5: Confusion matrix corresponding to the leave-one-out cross-validation testing.

In the second experiment we applied a cross-subject setting in which we divided the dataset in two halves with the actions of five different people each. The first half is used to train the model, the other half is used for testing it. We repeated this process 10 times and calculated the average and standard deviation for every action. The main goal of this experiment is to see how well our classifier generalizes when trying to classify new actions from previously unseen people. The classification results for this experiment are shown in Fig. 6, with an average correct classification rate of 68.63%. As we can see in the confusion matrix in Fig. 5 the classification results decrease when we transfer our classifier to new people. This makes sense since different people have different ways of carrying out the same activities. Furthermore, the number of samples used for the training is significantly inferior. This reduces the overall performances of the system especially for very similar actions.

Finally, our correct classification rates are comparable to the state of the art in action recognition where only body joints are considered [4]. However, our QTC_{3D} representation provides richer and more intuitive information about the spatio-temporal relation between those moving joints.

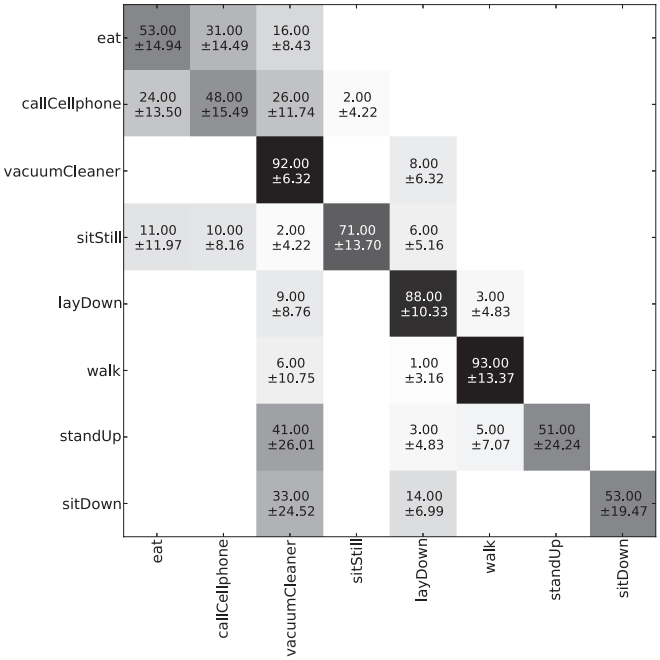


Fig. 6: Confusion matrix corresponding to the cross-subject testing.

VI. CONCLUSIONS

We presented a new application and practical implementation of QTC_{3D} to recognise human actions. Our correct classification rates are in accordance to state of the art in similar action recognition system based on depth data and human joints, showing that our QTC_{3D} approach is suitable for ADLs classification.

Future work includes extending our approach to select couples of joints that provide the most informative spatio-temporal relations for ADLs classification. Moreover, it would be interesting evaluate different generative models in order to determine the most suitable QTC_{3D} relations for describing specific activities.

The long term goal of this work is to develop a mobile robot for AAL that pro-actively searches for the best location to observe and successfully recognise ADLs in challenging domestic scenarios.

VII. ACKNOWLEDGEMENT

This work has been partly supported by H2020 EU Project ENRICHME (No: 643691, call: H2020-PHC-2014).

REFERENCES

[1] [Online]. Available: www.who.int
[2] [Online]. Available: www.oecd.org
[3] N. Mavridis, N. Bellotto, K. Iliopoulos, and N. V. de Weghe, "Qtc3d: Extending the qualitative trajectory calculus to three dimensions," *Information Sciences*, vol. 322, no. 0, pp. 20 – 30, 2015.
[4] J. Wang, Z. Liu, Y. Wu, and J. Yuan, "Learning actionlet ensemble for 3D human action recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 5, pp. 914–927, 2014.
[5] P. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea, "Machine recognition of human activities: A survey," *IEEE Transactions on Circuits and Systems for Video Technology,*, vol. 18, no. 11, pp. 1473–1488, 2008.
[6] J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, M. Cook, and R. Moore, "Real-time human pose recognition in parts from single depth images," *Communications of the ACM*, vol. 56, no. 1, pp. 116–124, 2013.
[7] A. G. Cohn and J. Renz, "Qualitative Spatial Representation and Reasoning," ser. Handbook of knowledge representation. Elsevier, 2008, vol. 3, pp. 551–596.
[8] M. Sridhar, A. Cohn, and D. Hogg, "Benchmarking qualitative spatial calculi for video activity analysis," in *IJCAI Workshop Benchmarks and Applications of Spatial Reasoning*, 2011, pp. 1–6.
[9] A. Behera, A. G. Cohn, and D. C. Hogg, *Workflow activity monitoring using dynamics of pair-wise qualitative spatial relations*. Springer, 2012.
[10] L. Santos, J. A. Prado, and J. Dias, "Human robot interaction studies on laban human movement analysis and dynamic background segmentation," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2009, pp. 4984–4989.
[11] C. Dondrup, N. Bellotto, and M. Hanheide, "A probabilistic model of human-robot spatial interaction using a qualitative trajectory calculus," *2014 AAAI Spring Symposium Series*, 2014.
[12] G. Pons-Moll, D. J. Fleet, and B. Rosenhahn, "Posebits for monocular human pose estimation," June 2014.
[13] J. Liu, B. Kuipers, and S. Savarese, "Recognizing human actions by attributes," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 3337–3344.
[14] H. Jhuang, J. Gall, S. Zuffi, C. Schmid, and M. J. Black, "Towards understanding action recognition," in *Proceedings of the 2013 IEEE International Conference on Computer Vision*, ser. ICCV '13. Washington, DC, USA: IEEE Computer Society, 2013, pp. 3192–3199.
[15] J. Wang, Z. Liu, J. Chorowski, Z. Chen, and Y. Wu, "Robust 3D action recognition with random occupancy patterns," in *European Conference on Computer Vision (ECCV)*, 2012.
[16] W. Li, Z. Zhang, and Z. Liu, "Action recognition based on a bag of 3d points," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshos (CVPR)*, 2010.
[17] O. Oreifej and Z. Liu, "Hon4d: Histogram of oriented 4d normals for activity recognition from depth sequences," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 716–723.
[18] X. Yang, C. Zhang, and Y. Tian, "Recognizing actions using depth motion maps-based histograms of oriented gradients," in *ACM international conference on Multimedia*, 2012, pp. 1057–1060.
[19] G. Chen, M. Giuliani, D. Clarke, A. Gaschler, and A. Knoll, "Action recognition using ensemble weighted multi-instance learning," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2014, pp. 4520–4525.
[20] X. Yang and Y. Tian, "Eigenjoints-based action recognition using naive-bayes-nearest-neighbor," in *IEEE Conference Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2012, pp. 14–19.
[21] G. Evangelidis, G. Singh, and R. Horaud, "Skeletal Quads : Human Action Recognition Using Joint Quadruples," *International Conference on Pattern Recognition (ICPR)*, pp. 4513–4518, 2014.
[22] L. Xia, C.-C. Chen, and J. Aggarwal, "View invariant human action recognition using histograms of 3d joints," in *The 2nd International Workshop on Human Activity Understanding from 3D Data (HAU3D) in conjunction with IEEE CVPR*, 2012.
[23] N. Van de Weghe, "Representing and reasoning about moving objects: A qualitative approach," Ph.D. dissertation, Ghent University, 2004.
[24] Y. Lu, "Discrete frenet frame with application to structural biology and kinematics," Ph.D. dissertation, Department of Mathematics, The Florida State University, 2013.
[25] S. M. LaValle, *Planning Algorithms*. Cambridge University Press, 2006.
[26] K. Iliopoulos, N. Bellotto, and N. Mavridis, "From sequence to trajectory and vice versa: Solving the inverse QTC problem and coping with real-world trajectories," in *AAAI Spring Symposium – Qualitative Representations for Robots*, 2014, pp. 57–64.
[27] P. Cimiano, A. Hotho, and S. Staab, "Comparing conceptual, divise and agglomerative clustering for learning taxonomies from text," in *Eureopean Conference on Artificial Intelligence (ECAI)*, 2004.
[28] M. Brand, N. Oliver, and A. Pentland, "Coupled hiddenMarkov models for complex action recognition," *IEEE Conference on Computer Vision and Pattern Recognition*, 1997.
[29] R. Bucks, D. Ashworth, G. Wilcock, and K. Siegfried, "Assessment of activities of daily living in dementia: development of the bristol activities of daily living scale," *Age and ageing*, vol. 25, no. 2, pp. 113–120, 1996.
[30] S. Katz, T. Downs, H. Cash, and R. Grotz, "Progress in development of the index of adl," *Gerontologist*, vol. 10, no. 1, pp. 20–30, 1970.