

Activity Recognition for Ergonomics Assessment of Industrial Tasks with Automatic Feature Selection

Adrien Malaisé, Pauline Maurice, Francis Colas, Serena Ivaldi

► **To cite this version:**

Adrien Malaisé, Pauline Maurice, Francis Colas, Serena Ivaldi. Activity Recognition for Ergonomics Assessment of Industrial Tasks with Automatic Feature Selection. IEEE Robotics and Automation Letters, IEEE 2019, 4 (2), pp.1132-1139. 10.1109/LRA.2019.2894389 . hal-01985013

HAL Id: hal-01985013

<https://hal.archives-ouvertes.fr/hal-01985013>

Submitted on 17 Jan 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Activity Recognition for Ergonomics Assessment of Industrial Tasks with Automatic Feature Selection

Adrien Malaisé¹, Pauline Maurice¹, Francis Colas¹ and Serena Ivaldi¹

Abstract—In industry, ergonomic assessment is currently performed manually based on the identification of postures and actions by experts. We aim at proposing a system for automatic ergonomic assessment based on activity recognition. In this paper, we define a taxonomy of activities, composed of four levels, compatible with items evaluated in standard ergonomic worksheets. The proposed taxonomy is applied to learn activity recognition models based on Hidden Markov Models. We also identify dedicated sets of features to be used as input of the recognition models so as to maximize the recognition performance for each level of our taxonomy. We compare three feature selection methods to obtain these subsets. Data from 13 participants performing a series of tasks mimicking industrial tasks are collected to train and test the recognition module. Results show that the selected subsets allow us to successfully infer ergonomically relevant postures and actions.

Index Terms—Recognition; Human Factors; Ergonomics; Feature Selection.

I. INTRODUCTION

WORK-related musculoskeletal disorders (WMSDs) are a major health issue, affecting almost 50% of industrial workers in developed countries, and representing an important cost for companies and society as a whole [1]. In order to reduce the prevalence of WMSDs, the ergonomics of the workplace should be evaluated and improved. Additionally, workers should be made aware of which movements and postures to avoid. Vignais et al. [2] indeed showed that real-time feedback about the ergonomics of a posture could help decrease the risk of developing WMSDs. Providing real-time ergonomic feedback to workers is one of the objectives of the European project AnDy [3]. Within the AnDy project, online ergonomic assessment will also be used to control a collaborative robot in a way that guides the worker towards more ergonomic movements.

Currently, ergonomics is evaluated with standard pen-and-paper worksheets filled by experts observing the workers doing their job [4]. Digital human modeling software provide automatic filling of these ergonomic worksheets [5], but only based on a digital simulation of the activity to evaluate. These software are intended for workstation design: they do not work

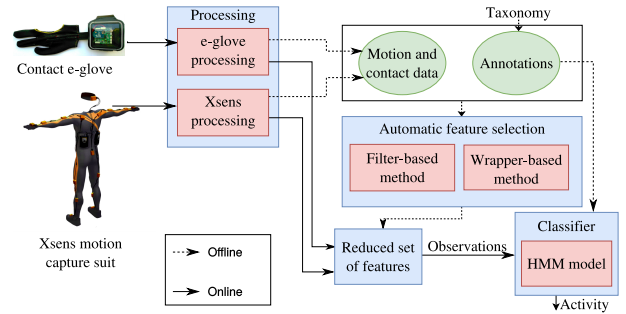


Fig. 1. Human activity recognition module. Data from both wearable sensors form the set of all available features, among which the feature selection module enables to select the best performing ones with either a filter-based method or a wrapper-based method. Parameters of the HMM models are learned based on the selected features and using an annotated training dataset (supervised learning). The feature selection and training processes are both performed offline. Recognition can be done either offline or online.

in real-time, and identification of actions performed is done manually by the user. A possible solution to automatize the filling of ergonomic worksheets is to use activity recognition algorithms to infer the actions and postures that are considered in these worksheets. In a previous paper [6], we proposed a method based on wearable sensors and Hidden Markov Models (HMM) to automatically recognize different activities during an industrial pick-and-place task. The activities were, however, high-level descriptions of the actions specific to the task considered and were not relevant for ergonomics evaluation. Furthermore, the features (i.e., task and movement descriptors, such as joint angles, center of mass position, etc.) used as input of the recognition model were manually selected among all the available data, which requires expert knowledge of the relevant movement descriptors for the tasks. To obtain robust and generic activity recognition models, one cannot rely on a small set of hand-picked features, because there is no guarantee that these features are the optimal set that maximizes the performance of the learned recognition models. On the other hand, using all the possible features is not a viable solution, since it requires to use a large number of wearable and environmental sensors, whereas in industrial applications the number of sensors should be optimized.

Motivated by these problems, in this paper we present two contributions to address the limitations of our previous work. First, we propose a generic taxonomy (i.e., a set of activities to recognize) that is compatible with the postures and actions evaluated in standard ergonomic assessment worksheets. Two ways of using the proposed taxonomy with activity recognition models based on supervised learning of an HMM are compared in order to identify the one that achieves the best

Manuscript received: September, 10, 2018; Revised December, 10, 2018; Accepted January, 08, 2019.

This paper was recommended for publication by Editor Allison M. Okamura upon evaluation of the Associate Editor and Reviewers' comments. This work was supported by the European Unions Horizon 2020 Research and Innovation Programme under Grant Agreement No. 731540 (project AnDy).

¹ Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France adrien.malaise@inria.fr

The authors wish to thank Clélie Amiot, Nicolas Paris and Guy Junior Richard from Université de Lorraine for the annotation of the database.

Digital Object Identifier (DOI): see top of this page.

recognition performance. Second, we propose to identify the best performing subsets of features used as input for our recognition algorithm. To be used in industrial environment, the number of required sensors for our activity recognition module must be minimized. Therefore, the most informative features regarding our proposed taxonomy need to be identified to select the optimal set of sensors. We compare different feature selection and dimensionality reduction methods in order to improve the robustness of the resulting list of relevant features.

We apply our proposed method to the recognition of activities relevant for ergonomics evaluation of industry tasks, that we recorded in a lab environment with 13 participants. The outcome of our study is the list of features to be used to classify the activities according to our taxonomy. Both the dataset and our software are also publicly available.

The paper is organized as follows: Section II reviews the related work. Section III presents the proposed taxonomy and the activity recognition module depicted in Fig. 1, and describes the creation of the dataset used to train and test the recognition module. The two ways of using the proposed taxonomy and the subsets of relevant features extracted from feature selection are compared and discussed in Section IV. Conclusions are presented in Section V.

II. RELATED WORK

This paper proposes to use an activity recognition algorithm to identify postures and actions considered in standard ergonomic assessment methods, in order to provide a basis for an automatic assessment of ergonomics. This section reviews how ergonomics is currently assessed in industry, and how such assessments could be automatized. Work related to recognition of human activities and feature selection is also presented.

A. Assessment of ergonomics in industry

Standard ergonomic assessment methods used in industry rely on pen-and-paper worksheets filled by an expert [7], *e.g.*, RULA (Rapid Upper Limb Assessment) [8], REBA (Rapid Entire Body Assessment) [9], OWAS (Owako Working Posture Analysis System) [10], OCRA (Occupational Repetitive Actions) [11] or EAWS (Ergonomic Assessment Worksheet) [12]. These ergonomic worksheets consists of grids or checklists that evaluate the main biomechanical factors of MSDs: posture, external force or manipulated load, and task repetitiveness. Posture is evaluated mainly based on the elevation of the arms and flexion of the torso.

Currently, no tools are used in industry to provide workers with real-time ergonomic information on their postures and gestures. Yet, Vignais et al. [2] showed that real-time ergonomic feedback could help reduce the factors of risk for WMSDs. Kim et al. [13] recently developed a wearable device with vibrotactile elements to warn workers about joint overloading. Their method focuses however only on load handling. Yet, awkward postures increase the risk of developing WMSDs even in the absence of a load. In order to assess the postural risk, motion capture systems can be used to track workers' whole-body motion in real-time and simultaneously fill in standard ergonomic assessment worksheets. Busch et al.

[14] used optical marker-based motion capture to fill in the REBA ergonomic assessment worksheet automatically and online. They however noted that marker-based motion capture systems are ill-adapted to industrial settings due to occlusion issues. Fully wearable sensors such as inertial measurement units (IMUs) can be a solution to the occlusion problem.

B. Automatic recognition of human activities

Aside from the motion tracking technique mentioned in the previous section, most methods proposed in the literature to automatically fill in ergonomic assessment worksheets rely on direct measurement of joints angles and segment positions [14][15]. Such methods enable to identify postures, but they cannot identify gestures or actions (*i.e.*, time-series of specific postures) which are present in some ergonomic assessment worksheets (*e.g.*, *walking* in EAWS worksheet). Moreover, action-related information can be needed to fill in the ergonomic worksheet (*e.g.*, is the person carrying a load, does the task involves vibrations...). Activity recognition models based on machine learning can be used to address the action recognition issue. Classification algorithms have been widely used to recognize human daily activities, such as walking, sitting or lying [16]. Among all the existing techniques, Hidden Markov Models (HMM) have the advantage of efficiently modeling time-series data such as whole-body human motions [17]. In a previous paper, we successfully used HMM models along with an inertial motion capture system to identify the phases of a pick-and-place activity [6].

C. Feature selection for activity recognition

Motion capture systems provide a large set of data that can be used as input of an activity recognition algorithm, *e.g.*, position, velocity and/or acceleration of all the joints and body segments. These features are however not all relevant: depending on the type of activities that are targeted and the classifier that is used, some features are useless or simply redundant. Relevant features can be selected without requiring any a priori knowledge by using feature selection algorithms. Feature selection algorithms allow to automatically select, from a high dimensional set of features, the subset that maximizes the recognition performance of a classifier, hence reducing the dimensionality of the input data. Two main types of methods exist for automatic feature selection: wrapper-based methods and filter-based methods [18].

1) *Wrapper-based methods*: With wrapper-based methods, the selection of relevant features is based on the recognition performance of a classifier and is thus classifier-dependent [19]. The general idea is overcome the combinatorial explosion by iteratively building a few best-performing candidates of feature subsets. More precisely, let N be the total number of features. First, the classifier is independently trained with each feature (*i.e.*, one model for each feature), and the recognition performance of each model is evaluated with standard metrics such as accuracy or F1-score. The M features that correspond to the best-performing models are retained for the next iteration, where M is a number chosen by the user. At the next iteration, a model is trained for each combination of one of








Standing (and walking)		Kneeling or crouching			
1		Standing & walking in alternation, standing with support	12		Upright
2		Standing, no body support (for other restrictions see Extra Points)	13		Bent forward
3		a	Bent forward (20-60°)	14	
		b	with suitable support		
4		a	Strongly bent forward (>60°)		
		b	with suitable support		

Fig. 2. Sample of the EAWS worksheet postural grid. Postures are first classified into main categories, such as standing or kneeling. Each main posture is then divided into subcategories corresponding to arms and torso configuration (e.g., bent forward, elbow at shoulder level).

the M retained features with any of the $N - 1$ other features. The M subsets of 2 features offering the best performance are then retained for the next iteration. The process is repeated for subsets of 3 features, and so on until the recognition performance plateaus. This iterative selection process has the advantage of taking into account relations between features, without requiring to test each possible subsets within the initial set of features. In the original method $M = 1$, *i.e.*, only the best subset of features is retained at each iteration. However, Mandery *et al.* [20] proposed to keep several subsets at each iteration to reduce the risk of eliminating a subset offering a high performance when later combined with one or more other features. They used the wrapper-based method with kinematic features of human motion, but each of their features was a high-dimensional vector. For instance, they considered all the joint angles of the human body as one single feature, hence not evaluating the relevance of each joint individually.

2) *Filter-based methods*: The selection of relevant features with filter-based methods is not classifier-dependent, therefore no time-consuming training of classifier is required. Each feature is individually assigned a score in relation with the observations. Interaction between features are not considered [21]. A score commonly used to rank the features in the Fisher score [22]: it depends on how this feature allows to regroup observations corresponding to the same activity (low intraclass variance), and differentiate observations that correspond to different activities (high interclass variance). Once each feature is assigned a score, the selection process consists in taking the highest-ranked features up to a given number.

III. METHOD

This section first presents the taxonomy and the recognition algorithm used to identify ergonomically critical postures and actions. The data collection campaign that was conducted to generate a dataset for training our models is then described.

A. Taxonomy

We base our work on the EAWS ergonomic assessment worksheet, as it is widely used in the industry. The taxonomy we propose, *i.e.*, the set of activities that need to be recognized by our activity recognition algorithm, is therefore inspired by the items evaluated in EAWS. The ergonomic assessment of EAWS is largely based on the evaluation of the posture. Nevertheless, posture is not the only risk factor for WMSDs,

and other factors evaluated in EAWS require contextual information about the action being performed (e.g., does the task involve vibrations, does the person carry a load). Our taxonomy is therefore composed of several levels to identify both postures and goal-oriented actions (Table I).

The postural assessment in EAWS identifies 4 main postures: standing, kneeling/crouching, sitting, lying. Each main posture is then divided into subcategories corresponding to arms and torso configuration: upright, bent, strongly bent, shoulder level work, and overhead work (see Fig. 2 for a sample of the EAWS postural grid). The labels in the postural level of our taxonomy (level DEPOS) mainly reproduce the EAWS postural classification, e.g., *standing bent forward*, *walking upright* (Table I). EAWS does, however, not distinguish between kneeling and crouching. These two postures may generate the same overall level of WMSDs risk, but is not necessarily located in the same joints or body-parts. In order to provide workers with joint-specific ergonomic feedback, our taxonomy distinguishes between kneeling and crouching.

The postural information contained in the DEPOS level of the taxonomy can be retrieved either directly, or by combining the identification of the main posture (standing, kneeling...) with the identification of the torso and arms configuration (bent, overhead work...). Separating the identification of the main posture from the identification of the torso and arms configuration allows to use a different recognition model for each of these two categories, which might improve the recognition performance. We therefore define 2 additional levels in the taxonomy: GEPOS and DET. GEPOS contains the labels corresponding to the main posture, while DET contains the labels corresponding to the torso and arms configuration. The two ways of inferring the full postural information is compared in this work to select the best-performing method (*i.e.*, using one single model with DEPOS labels or combining two models, one using GEPOS labels and one using DET labels).

The fourth and last level of the taxonomy (level CUACTION) contains labels representing goal-oriented actions, e.g., carrying a load, screwing. Unlike the postural levels of the taxonomy, the list of labels in the CUACTION level is not exhaustive. The labels described in Table I are sufficient to describe the actions performed in our experimental dataset (see Section III-C). New labels can however be added if other actions relevant for ergonomic assessment need to be identified.

B. Activity recognition with feature selection

In order to automatically infer the postures and actions proposed in our taxonomy, we use classifiers for human activity recognition which are described hereinafter. The software implementing our methods is available on Github (<http://github.com/inria-larsen/activity-recognition-prediction-wearable>).

1) *Classifier*: We build on our previous method [6], where activity models were represented by HMM with manually selected features, trained in a supervised way using action labels. Here, the activity models are also based on HMM, but with a set of features that is optimized by the feature selection algorithm. The models are trained in a supervised way to recognize the activities from the different levels of the taxonomy

TABLE I
DESCRIPTION OF THE TAXONOMY.

State	Label	Description
GEPOS		
Main posture		
Standing	St	Ends when feet start moving
Walking	Wa	Starts when one foot start moving, ends when both feet are still
Kneeling	Kn	At least one knee on the floor
Crouching	Cr	No knee on the floor
Sitting	Si	Buttock on a chair or support
Lying	Ly	Torso on the floor or horizontal surface
DET		
Torso and arms configuration		
Upright	U	Torso straight
Bent forward	BF	Torso flexion angle between 20° and 60°
Strongly bent forward	BS	Torso flexion angle greater than 60°
Shoulder level work	OS	Elbow(s) at or above shoulder level with hand(s) at or below head level
Overhead work	OH	Hand(s) above head level
DEPOS		
Full postural information: combination of GEPOS and DET		
	St_U	Standing upright
	St_BF	Standing bent forward
	St_BS	Standing strongly bent forward
	St_OS	Standing shoulder level work
	St_OH	Standing overhead work
	Wa_U	Walking upright

CUACT		
Goal-oriented action		
Reaching	Re	Moving an arm towards a target, no object in hand
Picking	Pi	Picking-up an object, starts when touching the object, ends when arm stops moving with respect to the body
Placing	Pl	Placing an object, similar to <i>Re</i> but with an object in hand
Release	Rl	Bringing arm back after manipulation task
Carrying	Ca	Carrying an object. Starts at the end of <i>Pi</i> , ends at the beginning of <i>Pl</i>
Fine manipulation	Fm	Dexterous manipulation of an object
Screwing	Sc	A special case of <i>Fm</i> : rotational screwing movement of the hand
Idle	Id	Not doing anything with hands

with the `hmmlearn` library (<http://hmmlearn.readthedocs.io/>) in Python. The models are defined by N states representing the activities, such as standing (St) or walking (Wa) for a model using the GEPOS level of the taxonomy (all activities are presented in Table I). $S = \{s_1, s_2, \dots, s_N\}$ represents the set of possible states. The model is trained based on several time series sequences, each represented by a series of discrete states $Q = \{q_0, q_1, \dots, q_t, \dots, q_T\}$, obtained by manual annotations (see Section III-C4), and a series of T observations $X = \{x_1, \dots, x_t, \dots, x_T\}$ corresponding to features data extracted from the motion capture suit or the glove. Section III-C2 presents the sequences of activity included in the dataset. For each instant t , the goal is to infer the activity, such as $q_t = s_i$.

To train and evaluate a model, the dataset is split randomly into a training set (70% of the sequences) and a testing set (30% of the sequences). Two parameters of the HMM are learned based on the data from the sequences of the training set: the modeled observations and the transition matrix. The

TABLE II
LIST OF FEATURES.

Group	Name	Dimensions
IMUs	orientation	17 * 4 = 68
	linear acceleration	17 * 3 = 51
	linear acceleration norm*	1
human model 3-DoF joints	angle	22 * 3 = 66
	angular velocity	22 * 3 = 66
	angular velocity norm*	1
	angular acceleration	22 * 3 = 66
	angular acceleration norm*	1
human model segments origin	3D position	23 * 3 = 69
	3D orientation	23 * 4 = 92
	3D velocity	23 * 3 = 69
	velocity norm*	1
	3D acceleration	23 * 3 = 69
	acceleration norm*	1
	3D angular velocity	23 * 3 = 69
	angular velocity norm*	1
	3D angular acceleration	23 * 3 = 69
	angular acceleration norm*	1
human model center of mass	3D position	3
	3D velocity	3
	velocity norm*	1
	3D acceleration	3
e-glove	acceleration norm*	1
	finger/palm pressure	4
	finger flexion	3

The “norm” features (marked with a *) correspond to the norm of the vector containing the velocity or acceleration of all joints or segments.

observations are modeled based on a Gaussian probability distribution, with a full covariance matrix used. A cross validation with ten iterations is computed in order to recognition performance for each model with the mean and standard deviation of the F1-score.

2) *Data processing*: To reduce noise, a sliding window filter is applied to the recorded motion capture and glove data. The observation vector contains the mean of the data across each time window. Banos et al. [23] showed that precise recognition could be obtained with short time windows (250 ms - 500 ms). A 250 ms window is used, with an overlap of 50% (125 ms) between each window, as it reduces the risk of missing data and can improve the performance [24]. Therefore, there is an observation every 125 ms.

3) *Feature selection*: The features represent the observation used to train the model. Table II presents the different features evaluated in our work, and Section III-C3 describes how they are obtained from our measurement devices. Both feature selection methods presented in Section II-C have advantages and drawbacks in terms of recognition performance and computation time. To optimize the choice of the relevant features to use as input of the classifier, three methods are used to extract relevant features from all possible features: a wrapper-based method, a filter-based method and a standard dimensionality reduction based on Principal Component Analysis (PCA). The resulting subsets of features are evaluated in terms of recognition performance.

As presented in Section II-C, the wrapper-based method is an iterative algorithm. The method used is inspired by the work of Mandery et al. [20]. The HMM models presented above are used as classifier to evaluate the subsets of features, and the

metric performance used is the F1-score. At each iteration k , the 10 best subsets of features of dimension k are kept to be tested at iteration $k + 1$. The filter-based method used is based on the Fisher score with regard to II-C2. For each subset of features obtained with the three different methods, a cross-validation of three iterations is performed with the dataset split randomly at each iteration with 70% of the sequences used to train the model, and 30% to test the model.

C. Experimental protocol

The activity recognition algorithm presented in section III-B is based on supervised learning. Thus, a labeled dataset must be provided for the training stage. We conducted a campaign of in-lab data collection where participants performed various manual tasks. Whole-body kinematics and hand contact forces were recorded. The data were then manually annotated to generate the training dataset. The dataset is publicly available on Zenodo (<http://zenodo.org/record/1471975>). The experiment was approved by Inria's ethical committee COERLE.

1) *Participants*: 13 healthy adults took part in the data collection (9 males and 4 females, 11 right handed and 2 left handed, age: 25.7 ± 5.0 yrs, body height: 175.4 ± 7.9 cm, body mass: 72.3 ± 14.4 kg). Participants were naive to the purpose of the experiment and had no or limited experience of industrial work. All participants gave written informed consent before starting the experiment.

2) *Task description*: Our activity recognition software specifically targets tasks and postures that are common in industrial environments, following the taxonomy proposed in section III-A. Therefore, participants were asked to perform a series of manual activities mimicking industrial activities and involving postures of the EAWS ergonomic worksheet. Six activities were defined and are depicted in Fig. 3:

- *Screw high*: Take a screw and a bolt on a 75 cm-high table, walk to the shelf, screw at a height of 175 cm.
- *Screw middle*: Take a screw and a bolt on a 75 cm-high table, walk to the shelf, screw at a height of 115 cm.
- *Screw low*: Take a screw and a bolt on a 75 cm-high table, walk to the shelf, screw at a height of 25 cm (6 participants) or 60 cm (7 participants).
- *Untie knot*: Untie a knot placed on a 45 cm-high table.
- *Carry 10 kg*: Take a 10 kg load on a 55 cm-high table, walk to the shelf, put the load on a 20 cm-high shelf.
- *Carry 5 kg*: Take a 5 kg load on a 55 cm-high table, walk to the shelf, put the load on a 110 cm-high shelf.

One trial consisted in performing all 6 activities successively in a specified order. The order in which the activities had to be performed was selected randomly among a list of 6 predefined orders. Each participant performed the activities in 3 different orders, with 5 consecutive trials for each order, resulting in a total of 15 trials. Each trial lasted approximately 90 s. A break of about 2 min was allowed between each trial to limit fatigue. Two spatial organizations of the items involved in the activities were used to avoid any direct association of one activity with one position in space (set-up A: 6 participants; set-up B: 7 participants). Participants were given no instructions regarding the way to perform each activity. The positions of the different

items (holes and tables heights, position of the knot on the table) were, however, chosen to encourage participants to adopt some specific EAWS postures (*e.g.*, crouching, working with hand above head level...).

3) *Instrumentation*: Two kinds of wearable sensors were used to record whole-body motion and contact forces (Fig. 1). Participants were equipped with an Xsens MVN Link inertial motion tracking suit consisting of 17 IMUs to record whole-body kinematics (Xsens Technologies, Enschede, The Netherlands). Kinematic data were recorded at 240 Hz with the Xsens MVN software. The MVN software outputs the orientation and linear acceleration of each IMU. In addition, the MVN software uses a scalable kinematic body model composed of 23 segments linked together by 22 3-DoF joints to compute the position, velocity and acceleration of the origin of each segment and of the center of mass of the model. Participants were also equipped with an instrumented e-glove on their right-hand (Emphasis Telematics, Athens, Greece). The e-glove embeds 3 flexion sensors – on the thumb, index, and middle finger – and 4 pressure sensors – on the palm and on the fingertip of the thumb, index and middle finger. Data of the e-glove were recorded at 50 Hz with a proprietary software. The Xsens and e-glove data were timestamped to enable their synchronization. All Xsens and e-glove outputs were integrated in the features used in the activity recognition algorithm (Table II). In addition to these local features concerning individual joint, we used “norm” features that correspond to the norm of the vector containing the velocity or acceleration of all joints or segments. These features are inspired from Mandery *et al.* [20] who identified them as relevant for their activity recognition. Note that no calibration of the e-glove signal was performed: only the raw sensors data were used. In addition to the wearable sensors, participants were videotaped with 2 cameras. Data of the cameras were not used in the activity recognition algorithm. They only served to help annotators resolve ambiguities when labeling the training dataset.

4) *Annotations*: In order to generate a training dataset, the collected data must be manually labeled using the labels defined in Table I. Labeling by human annotators is intrinsically subjective, especially during transitions between two successive postures or tasks. Therefore three annotators independently labeled the data. The ground truth is extracted from these annotations. For each time window of 250 ms (see Section III-B2) there are six related frames and annotations per annotators. Thus, 18 labels are associated to a window of observations. The ground truth is defined by the label which had the majority among the 18 labels. An absence of majority can happen during transition or when an annotator adds an intermediate state (*e.g.*, a bent before a strongly bent). To resolve this conflict, the segments without majority are cut with the first half labeled as the label of the previous state, and the second half labeled as the label of the next state. Annotations were performed with the Anvil software (<http://www.anvil-software.org/>) on the videos of the Xsens avatar replaying participants' motions (video framerate: 25 fps). Each frame of the video was given 4 different labels: one for each level of the taxonomy defined in Section III-A.

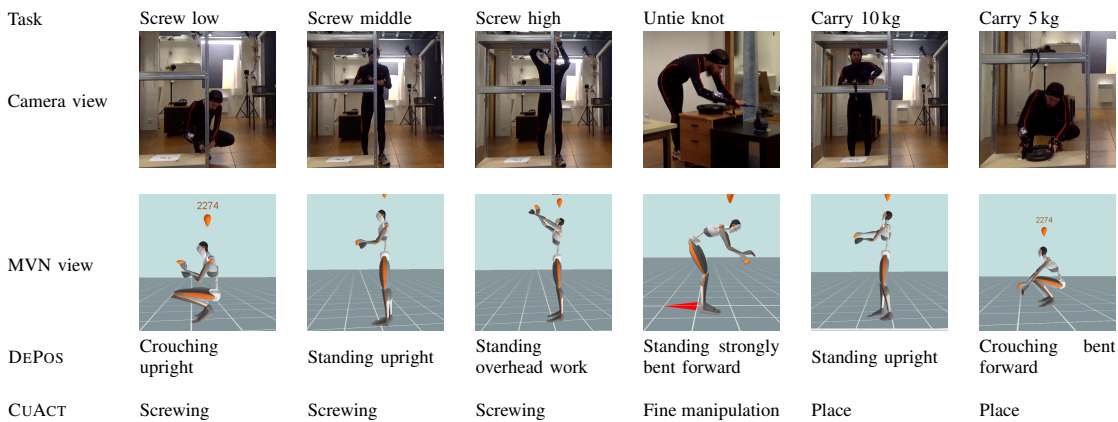


Fig. 3. The 6 tasks performed by participants, with associated videos and corresponding posture and action following the taxonomy proposed in Section III-A.

The annotators give labels corresponding to the level: GEPOS, DEPOS and CUACT. The DET is directly extracted from the DEPOS labels, by keeping only the labels about the additional postural information (bent, overhead work...).

IV. RESULTS

First, the performance of the subsets of features identified with the 3 feature selection methods are compared. Then these taxonomy-specific subsets are compared with generic sets of features. A third part compares the use of a direct model and a combined model to represent the taxonomy. The last part presents the results of generalization between participants.

A. Comparison of the feature selection methods

Fig. 4 displays the evolution of the recognition performance (F1-score) with the number of features included in the recognition model, for all 3 feature selection methods (wrapper-based, filter-based and PCA) and for each of the 4 taxonomy levels. Table III then presents the dimension and recognition performance of the best subsets of features identified with each of the 3 selection methods. In both cases, we limited the size of the subsets to 15 features, since we aim to use low-dimensional sets of features to increase computational efficiency and reduce the number of sensors needed. For all 4 levels of the taxonomy, the wrapper-based method identifies better performing subsets than both the filter-based method and the PCA. The recognition scores are nevertheless close with all 3 methods, except for the CUACT level where the wrapper-based method set largely outperforms the 2 other sets of features. Importantly when aiming to reduce the number of sensors, the wrapper-based method identifies well-performing sets of features even in small dimension: the recognition performance starts to plateau around 3 features (Fig. 4). It should also be noted that though subsets of features identified with the simple PCA method give good recognition performances in low dimension, they cannot be used to reduce the number of sensors. Outputs of a PCA are linear combinations of the input data. Therefore all initial features (hence all sensors) are needed to compute the features selected with the PCA.

Fig. 4 shows that adding new features to subsets selected with the filter-based method sometimes degrades the recognition performance (e.g., for DEPOS and DET). This is likely

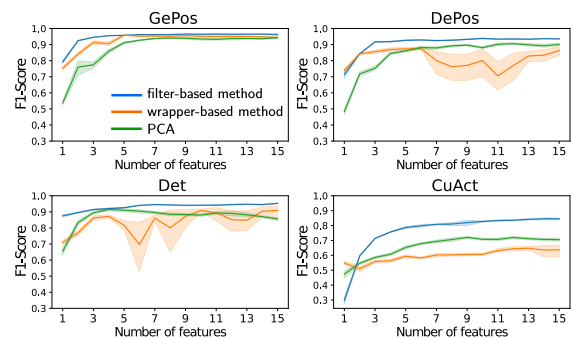


Fig. 4. Evolution of F1-score with the number of features used for the recognition. Full lines and shaded areas represent respectively the mean of the F1-Score and the standard deviation. The wrapper-based method allows to reach the maximum of performance with a low number of features. With the filter-based method, addition of certain features can decrease the performance such as for the DET and DEPOS levels.

due to the underlying Gaussian assumption in the observation model of the HMM: if the distribution of the data related to one feature strongly differs from a Gaussian distribution, inclusion of this feature in the model might deteriorate the performance. The wrapper-based method is less sensitive to the violation of the Gaussian assumption in features, because features are then selected directly based on the recognition performance of the model. Hence the wrapper-based method is more robust, but its feature selection process is computationally much more expensive than the filter-based and PCA methods. This last concern nevertheless fades if optimal sets of features associated with the proposed taxonomy are identified once and for all, and the selection process does not need to be run again.

B. Comparison of subsets of features

In order to check whether identifying sets of features specific to the proposed taxonomy is needed, we compared our best-performing subsets of features (with the wrapper-based method) with subsets proposed by Mandery et al. [20] for human activity recognition, and with high-dimensional sets of features (all features, 3D position of each segment, and angle of each joint). Since the 2 subsets proposed in Mandery et al. [20] are of dimension 4 and 8, we used our best-performing subsets of similar dimension to make

TABLE III
PERFORMANCE OF THE FEATURE SELECTION METHODS

Taxonomy		Wrapper	Filter	PCA
GEPOS	Dimensions	9	5	15
	F1-score	96.53	95.81	94.28
DET	Dimensions	15	15	4
	F1-score	95.25	90.90	91.50
DEPOS	Dimensions	10	6	12
	F1-score	93.81	87.58	90.55
CUACT	Dimensions	15	13	9
	F1-score	84.55	64.77	72.06

the results comparable. Table IV presents the recognition performance obtained with those different sets of features. First, using high-dimensional sets of features does not result in better performance compared to using reduced sets of selected features, while it is likely computationally less efficient and may require more sensors. Hence smaller sets of features are advantageous. Second, among the reduced sets of features, our best-performing subsets greatly outperform the generic subsets proposed by Mandery *et al.* [20] for all 4 levels of the taxonomy. Hence, using a set of features dedicated to our taxonomy is recommended in order to optimize the performance of the recognition module. It should nevertheless be noted that the scores presented here are obtained with subsets of features optimized separately for each level of the taxonomy. If all levels are to be used simultaneously, and if a goal is to reduce the number of sensors required, a subset of features common to all 4 levels may be preferable.

Table V lists the subsets of features that allows to reach 2% of the maximum recognition rate for each taxonomy level (See Table III). The GEPOS and DET subsets both contain only features that can be obtained with a reduced number of sensors. GEPOS features correspond to lower-body movements, while DET features are related to back and arm movements, in accordance with the labels of each of these 2 taxonomy levels (Table I). Conversely, several features in the DEPOS and CUACT subsets require all IMUs sensors for their computation (center of mass and “norm” features). These “aggregated” features were included in the initial set because Mandery *et al.* [20] identified them as useful for human motion recognition which is confirmed by our results. Nevertheless, when the number of sensors is a concern, a new set of features should be identified with excluding all “aggregated” features from the initial set of available features.

C. Comparison of models associated with the taxonomy

In the remaining of the paper, we only use the subsets presenting in Table V. We compare two ways of inferring the full postural information: using one model to recognize directly the full posture (DEPOS) or combining two models, each one focusing on different aspects of the posture (COPOS = GEPOS + DET). The two methods are evaluated by computing the F1-score of each sequence obtained either with the direct model or with the combined model.

Using only a single model to infer directly DEPOS gives a better recognition performance: the F1-score is 92.56 with

TABLE IV
PERFORMANCE OF DIFFERENT SUBSETS OF FEATURES

Features	Dim.	GEPOS F1	DET F1	DEPOS F1	CUACT F1
<i>Our subsets</i>					
Best subset dimension 4*	4	95.75	92.01	91.84	75.65
Best subset dimension 8*	8	96.19	94.30	92.72	81.00
<i>Mandery et al. [20]</i>					
Center of mass velocity, Segments velocity norm	4	88.61	32.10	30.63	23.75
Center of mass velocity, Segments velocity norm, Velocity norm of hands and feet	8	48.59	82.35	67.22	55.12
<i>High-dimensional</i>					
All available features	779	93.95	93.16	89.30	81.39
All segment positions	69	93.34	95.62	91.15	71.56
All joint angles	66	91.37	93.74	90.12	73.51

TABLE V
DETAILS OF THE FINAL SELECTED SUBSETS OF FEATURES.

Taxonomy	Dim.	F1-score	Subsets of features
GEPOS	3	94.53	Pelvis position (z) Pelvis velocity (z) Left lower leg acceleration (x)
DET	6	94.00	Angular acceleration pelvis (y) L5 vertebra orientation (q1, q3) Right shoulder orientation (q2) Left forearm position (z) Right hand position (x)
DEPOS	5	92.61	Center of mass position (z) Segments velocity norm Right shoulder orientation (q1) Right ankle joint angle (z) Right upper leg position (z)
CUACT	11	83.36	Segments velocity norm IMU acceleration norm Angular acceleration head (z) Right elbow joint angle (z) Pelvis orientation (q3) Left hand position (x) Left hand orientation (q3) Head velocity (z) Neck velocity (z) Right forearm velocity (z) Right hand velocity (z)

the DEPOS model, and 89.68 with the combined COPOS model. The DEPOS model relies on a lower number of features (5) than the combined model (9 features since no features are shared between the GEPOS and DET subsets), which may render the DEPOS model computationally more efficient. However, as mentioned in Section IV-B the features in the DEPOS subsets require all input data, hence all sensors. The DEPOS model is not more efficient than the COPOS regarding the number of sensors required. The combined model can also be useful to infer full postures that are not in the dataset, if the corresponding labels of the GEPOS and DET levels are included in the dataset. For example, if the real state is *kneeling bent forward* and does not appear in the dataset, the combined model could infer the complete posture if both *kneeling* and *bent forward* states are known individually.

D. Generalization

The recognition models should be able to recognize activities of users that are not in the training set. We therefore test the generalization capability of the recognition models with a leave-one-out cross validation. For each of the 13 participants, a model is trained on the data of the 12 other participants, and the recognition performance is tested with the data of the participant excluded from the training set (the best performing sets of features identified with the wrapper-based method are used). To be consistent with the previous section, only 70% of the sequences of the 12 participants in the training set are used. The F1-score is computed for each of the 13 validations, and the global evaluation of the generalization corresponds to the average score of the 13 F1-scores. The global F1-score is 94.00 for the GEPOS model (2.62% of max. F1-score), 91.55 for the DET model (3.88% of max. F1-score), 90.96 for the DEPOS model (3.04% of max. F1-score), and 81.91 for the CUACT model (3.12% of max. F1-score). The proposed subsets of features for each taxonomy level show overall a good generalization performance.

V. CONCLUSION

In this paper, we addressed the problem of automatic assessment of ergonomics. We proposed a taxonomy of postures and actions inspired by the EAWS ergonomics assessment worksheet used in industry. We showed that using activity recognition models based on HMM with automatic feature selection, postures can be successfully recognized with only a small number of features. In addition, recognition is better when using subsets of features identified specifically for our taxonomy than when using generic features. Therefore, we provide the list of relevant subsets of features associated with our taxonomy. Recognition of actions did not exhibit similar performance, and might require a more complex model such as one based on Neural Network. In the future, the activity recognition module will be used to provide on-line ergonomic feedback to warn workers of non-ergonomic movements.

REFERENCES

- [1] E. Schneider, X. Irastorza, M. Bakhuis Roozeboom, and I. Houtman, "Osh in figures: occupational safety and health in the transport sector-an overview," *EU-OSHA*, 2010.
- [2] N. Vignais, M. Miezal, G. Bleser, K. Mura, D. Gorecky, and F. Marin, "Innovative system for real-time ergonomic feedback in industrial manufacturing," *Applied ergonomics*, vol. 44, no. 4, pp. 566–574, 2013.
- [3] S. Ivaldi, L. Fritzsche, J. Babic, F. Stulp, M. Damsgaard, B. Graitmann, G. Bellusci, and F. Nori, "Anticipatory models of human movements and dynamics: the roadmap of the andy project," in *DHM*, 2017.
- [4] G. Li and P. Buckle, "Current techniques for assessing physical exposure to work-related musculoskeletal risks, with emphasis on posture-based methods," *Ergonomics*, vol. 42, no. 5, pp. 674–695, 1999.
- [5] T. Bossomaier, A. G. Bruzzone, A. Cimino, F. Longo, and G. Mirabelli, "Scientific approaches for the industrial workstations ergonomic design: A review." in *ECMS*, 2010, pp. 189–199.
- [6] A. Malaisé, P. Maurice, F. Colas, F. Charpillat, and S. Ivaldi, "Activity recognition with multiple wearable sensors for industrial applications," in *ACHI*, 2018.
- [7] D. Roman-Liu, "Comparison of concepts in easy-to-use methods for msd risk assessment," *Applied ergonomics*, vol. 45, no. 3, pp. 420–427, 2014.
- [8] L. McAtamney and E. N. Corlett, "Rula: a survey method for the investigation of work-related upper limb disorders," *Applied ergonomics*, vol. 24, pp. 91–99, 1993.
- [9] S. Hignett and L. McAtamney, "Rapid entire body assessment," in *Handbook of Human Factors and Ergonomics Methods*. CRC Press, 2004, pp. 97–108.
- [10] E. N. Corlett and R. Bishop, "A technique for assessing postural discomfort," *Ergonomics*, vol. 19, no. 2, pp. 175–182, 1976.
- [11] E. Occhipinti, "Ocr: a concise index for the assessment of exposure to repetitive movements of the upper limbs," *Ergonomics*, vol. 41, no. 9, pp. 1290–1311, 1998.
- [12] K. Schaub, G. Caragnano, B. Britzke, and R. Bruder, "The european assembly worksheet," *TIES*, vol. 14, no. 6, pp. 616–639, 2013.
- [13] W. Kim, M. Lorenzini, K. Kapicioglu, and A. Ajoudani, "Ergotac: A tactile feedback interface for improving human ergonomics in workplaces," *IEEE RA-L*, 2018.
- [14] B. Busch, G. Maeda, Y. Mollard, M. Demangeat, and M. Lopes, "Postural optimization for an ergonomic human-robot interaction," in *IROS*. pp. 2778–2785, 2017
- [15] P. Plantard, H. P. Shum, A.-S. Le Pierres, and F. Multon, "Validation of an ergonomic assessment method using kinect data in real workplace conditions," *Applied ergonomics*, vol. 65, pp. 562–569, 2017.
- [16] A. Dubois and F. Charpillat, "Human activities recognition with RGB-Depth camera using HMM," in *EMBC*, 2013, pp. 4666–4669.
- [17] J. Yamato, J. Ohya, and K. Ishii, "Recognizing human action in time-sequential images using hidden markov model," in *CVPR*, 1992, pp. 379–385.
- [18] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of machine learning research*, vol. 3, no. Mar, pp. 1157–1182, 2003.
- [19] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial intelligence*, vol. 97, no. 1-2, pp. 273–324, 1997.
- [20] C. Mandery, M. Plappert, J. Borrás, and T. Asfour, "Dimensionality reduction for whole-body human motion recognition," in *FUSION*, pp. 355–362, 2016.
- [21] J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, and H. Liu, "Feature selection: A data perspective," *ACM CSUR*, vol. 50, no. 6, p. 94, 2017.
- [22] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. John Wiley & Sons, 2012.
- [23] O. Banos, J.-M. Galvez, M. Damas, H. Pomares, and I. Rojas, "Window size impact in human activity recognition," *Sensors*, vol. 14, no. 4, pp. 6474–6499, 2014.
- [24] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," in *PerCom*, pp. 1–17, 2004.