FOCUS

Human activity recognition based on a sensor weighting hierarchical classifier

**Oresti Banos · Miguel Damas · Hector Pomares ·
Fernando Rojas · Blanca Delgado-Marquez ·
Olga Valenzuela**

Abstract The analysis of daily living human behavior has proven to be of key importance to prevent unhealthy habits. The diversity of activities and the individuals' particular execution style determine that several sources of information are normally required. One of the main issues is to optimally combine them to guarantee performance, scalability and robustness. In this paper we present a fusion classification methodology which takes into account the potential of the individual decisions yielded at both activity and sensor classification levels. Particularly tested on a wearable sensors based system, the method reinforces the idea that some parts of the body (i.e., sensors) may be specially informative for the recognition of each particular activity, thus supporting the ranking of the decisions provided by each associated sensor decision entity. Our method systematically outperforms the results obtained by traditional multiclass models which otherwise may require a high-dimensional feature space to acquire a similar performance. The comparison with other activity-recognition fusion approaches also demonstrates our model scales significantly better for small sensor networks.

O. Banos (✉) · M. Damas · H. Pomares · F. Rojas
Department of Computer Architecture and Computer
Technology, CITIC-UGR, University of Granada,
C/Periodista Daniel Saucedo Aranda s/n,
18071 Granada, Spain
e-mail: oresti@atc.ugr.es

B. Delgado-Marquez
Department of International Economics,
University of Granada, Granada, Spain

O. Valenzuela
Department of Applied Mathematics,
University of Granada, Granada, Spain

Keywords Multisource fusion · Hierarchical classification · Weighted decision · Binary classifiers · Activity recognition · Wearable sensors

1 Introduction

The enhancement of the quality of life of the elderly is one of the most important goals considered in the ambient assisted living framework (European-Commission 2011). For that purpose, the main idea is to reduce the innovation barriers of forthcoming promising markets, and to lower future social security costs through the use of the potential offered by the information and communication technologies (ICTs). The motivation of this new funding activity is in the demographic change and ageing in the developed countries, which implies not only challenges but also opportunities for the citizens, the social and healthcare systems as well as the industry or the market.

One of the main premises to guarantee an adequate quality of life is to define proactive policies. The idea is not just to act when the subject requires to be treated but to avoid such situations as far as possible. Since the eventual health problems might be tackled before coming true, it is likely the subject will not experience the disease and consequently not require future assistance. This may help to reduce the costs required to afford the treatments, which in some cases are particularly expensive. Obviously some damages associated with uncontrollable factors may not be averted even considering proactive conducts. Nevertheless, some of the most relevant may be prevented. As an example, cardiovascular diseases, cancer or diabetes lead causes of death and disability in the United States accounting for 60–70 % of all deaths in the Region (WHO 2006). These diseases share common risk factors which

include tobacco use, physical inactivity, obesity and hypertension. There is sufficient evidence these diseases can be prevented and controlled through changes in lifestyle, public policies and health interventions. If these risk factors were eliminated, at least 80 % of all heart diseases, strokes and type-2 diabetes and over 40 % of cancer cases could be prevented (WHO 2005). Undoubtedly, this is an encouraging example which reinforces the need of the integration of preventive methods to guarantee a better quality of life and a sustainable healthcare system.

The analysis of the daily living subjects' behavior is of key importance to identify possible unhealthy conducts. The traditional methods require the direct intervention of the users, normally filling out periodic questionnaires and reports. One of the main problems of this kind of techniques is the memory limitations of individuals and intentional misreport. An increasing lack of interest in the report task is experimented by the user due to the required continuous involvement. These issues along with problems with reliability, validity and sensitivity have been comprehensively summarized in Shephard (2003).

Nevertheless, these problems may be faced with the use of systems which perform the monitoring task without the users' participation. The latest advances on the sensor monitoring technologies allow us to define a new generation of systems which may automatically and autonomously perform the recordings. For example, the assessment of tobacco use (Sazonov et al. 2011), digestive disorders (Sazonov et al. 2010) or the sedentariness (Staudenmayer et al. 2009; Bonomi et al. 2009) may be performed without direct intervention of the user.

A detailed description of the subject behavior may help to have a better understanding of the eventual health problems the individual might suffer. Nevertheless, the dimensionality of the context and activity-recognition problems is sometimes misunderstood and some of the proposed methods just succeed in restricted scenarios (Warren et al. 2010). Because of that the current tendency is to define sensor networks on and around the subject which take into consideration a broad representation of the individuals' behavior and their surroundings (Roggen et al. 2010). This is supported by the idea that decisions based on the collectivity are required to deal with the complexity of the activity-recognition problem. Nonetheless, the proposed fusion models generally lack of scalability and/or reliability capabilities since the same importance is usually given to the decisions provided by all the sources. This is particularly harmful when small sensor networks are considered. Here, we are on the definition of a fully scalable fusion method which supports the efficient combination of different sources of information through considering the particular discriminatory potential of each one. The idea is further extended to the activity or class level, since each

source may be particularly suitable for the recognition of a subset of the whole set of target activities.

The rest of the paper is organized as follows. In Sect. 2, a brief summary of the main activity-recognition issues and the classical methodology is introduced. Section 3 presents the most common machine-learning techniques used in activity recognition and the proposed fusion method. In Sect. 4, the considered models are tested for a particular case. The results are subsequently discussed in Sect. 5 and our final conclusions are summarized in Sect. 6.

2 Activity-recognition issues

2.1 Sensor modalities

The complexity of human daily living activity recognition resides on the diversity of possible executions and context situations that may refer to the same activity. Since many of them are related to the body movement (or absence), the use of motion sensors has become one of the most recurrent alternatives in the literature.

For the motion analysis, two types of sensors are generally used depending on whether they are placed on (wearable sensors) or around (ambient sensors) the subject. In principle, the use of ambient sensors such as cameras or microphones is restricted to particular scenarios where their deployment is feasible. Even when practicable, there exist some additional constraints that may difficult their use (subject's privacy, occlusions, ambient noise, etc.). Consequently there is an increasing tendency to the use of on-body sensors lacking such kind of limitations.

From the range of sensors that may be attached to the body, the inertial sensors are the most exhaustively used, providing good results for different setups and activities. Nevertheless, one of the main drawbacks is the obtrusiveness. Several considerable-sized devices attached to different parts of the subject's body in a likely uncomfortable way may not be considered a realistic daily wearable solution. Fortunately sensors miniaturization, battery-consumption reduction and lower cost production allow us to envision a new generation of tiny sensors to be integrated in "wearable things" such as the subject's clothes (Amft and Lukowicz 2009).

2.2 Exogenous factors

There are some well-known characteristics of the main daily activities that in theory allow us to accurately discriminate them. For example, depending on the intensity of the movements, one should be able to distinguish between some exercises such as walking and running, which nevertheless share a common execution style. The orientation

of the body may provide useful information about the posture when carrying out low intensity or quasi-static activities (e.g., lying down vs. standing still). However, there exist exogenous factors that difficult the recognition task. Age, weight, height or other subject-related features, as well as ambient and context-related factors (e.g., subject carrying items, unstable floor, etc.) may determine that notably different data recordings could refer to a similar activity. As an example, one cannot expect to register the same kind of data when an adult is cycling as when an elder does; similarly, the gait may differ when walking on the ground, grass or a frozen surface.

2.3 Activity-recognition chain

To deal with most of these problems, a general methodology is proposed. Classically referred to as activity-recognition chain (ARC), a general scheme is shown in Fig. 1. From left to right, a set of M sources (sensors) delivers raw unprocessed signals (u_j) representing the magnitude measured (e.g., acceleration). The signals are usually preprocessed (p_j) to avoid noise and diverse nature artifacts, typically through a filtering process. To capture the dynamics of the signals, these are partitioned in segments of a certain length (s_{jk}). Different techniques are devised for that purpose, mainly based on windowing or event-activity-based segmentation. Subsequently a feature-extraction process is carried out to provide a handy representation of the signals for the pattern-recognition stage. A wide range of heuristics, time/frequency domain and other sophisticated mathematical and statistical functions are commonly used. The feature vector is provided as input of the classifier ($f(s_{jk})$),

which ultimately assigns the activity or class recognized (c_i) to one of the N considered for the particular problem. The last optional stage corresponds to a fusion model that may combine the decisions of each individual ARC to improve the reliability of the recognition system. An extensive topical review of the main stages of the ARC may be seen in Preece et al. (2009b).

Each stage of the ARC may suffer from different kinds of difficulties. At the signal level, and beyond the special characteristics of each sensor, the data may contain artifacts and noise. Depending on the type of signals and target activities the classical preprocessing techniques may be more or less appropriate. For example, some filtering processes may imply an information loss. Since the eventual removed data can nevertheless be informative for some of the considered activities, the selection of an adequate preprocessing technique should be carefully studied. With respect to the segmentation process, it is not clearly defined which window data length must be considered. In general it depends on the complexity, duration and granularity of the activities among other considerations. The feature extraction process usually constitutes the computational bottleneck of the ARC. Since the final goal may be to define a real-time activity recognition system it is important to look for features which are not too much costly in terms of resources or reduce the number of features required to the minimum possible (efficiency optimization). Such reduction may in general help to define simpler classifiers. The other problem is the search of the optimal feature vector (performance optimization). Ideally, all the possible combinations of the extracted features should be assessed, implying an exponential complexity search problem (O^n).
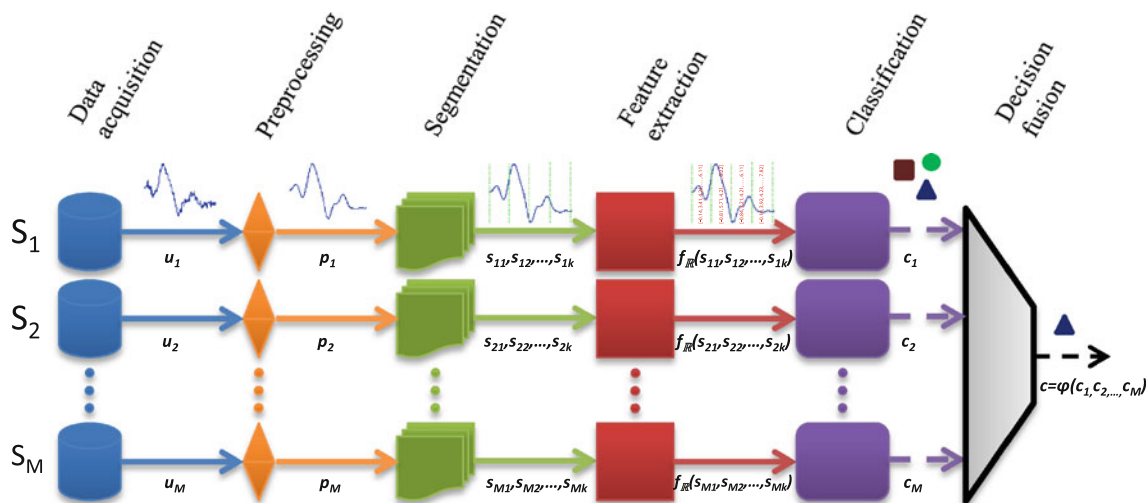
Fig. 1 Multiple activity recognition chain (M-ARC). M sensors deliver raw signals (u_j) which are subsequently processed (p_j). The signals are k-partitioned (s_{jk}) and a set of features (generically defined as f) are extracted from them, possibly different for each chain. The feature vector is used as input to the classifier entities. Each classifier yields a class on a N-class problem which may be combined through a fusion decision method. The indices are respectively defined for all $j = 1, \ldots, M; k = 1, \ldots, K; i = 1, \ldots, N$.

Filter feature-selection methods allow us to reduce the search, but in general, they do not ensure the best subset of features to be selected. On the other hand, wrapper methods evaluate the features' capabilities when used on the classifiers, thereby providing a more reliable representation of the feature potential but likely requiring a huge amount of computational resources and time. On top of the ARC, machine-learning methods are more specifically affected by problems related to their practical use, normally unbalanced data or boundaries convergence. The generality of the ARC definition determines all these problems may be faced in a different way depending on the considered sensor modalities. However, not all the solutions are in line with the definition of a usable model with independence of the particular context and setup.

Even whether the ARC is optimized, it is clear that some sensors may be more specialized in the recognition of some activities. Consequently, to increase the recognition scope, a combination of sensors is suggested. This combination or fusion may be performed at each level of the ARC (Sharma et al. 1998). For example, the sensor fusion may be performed at the feature-extraction level, thus defining a single feature vector composed by the independent features extracted from each sensor (Mantyjarvi et al. 2001). The arising problem is that one might have to deal with a high-dimensional feature space problem, which may difficult the feature selection and classification process as described above. This approach is not scalable for an online learning, since the inclusion of a new sensor requires to redefine the feature vector and retrain the whole system.

More interesting may be to define a fusion scheme acting at the classification level (Fig. 1). The idea is basically to combine the decisions delivered by each individual ARC in one single reinforced decision. In general, the system based on the fusion will be more accurate and robust than the individual ARCs. Furthermore, unlike (the) fusion at the feature level, the addition or removal of sensors would not imply to retrain the initial systems but to update the decision structure parameters. The main problem of the decision fusion is to define an accurate model which takes into account robustness and scalability. That means the model should be accurate enough with independence of the topology or the number of sensors considered.

3 Classification methods for activity recognition

3.1 Direct multiclass models

As Fig. 1 shows, a single ARC may be defined for each source or sensor ideally constituting a complete autonomous recognition system. Different classification schemes may be used on top of the single ARCs.

Decision trees (DT) proved to perform well in combination with time and frequency domain features (Bao and Intille 2004; Maurer et al. 2006; Parkka et al. 2006) although they were less accurate for other setups (Ermes et al. 2008). DT algorithms examine the discriminatory ability of the features one at a time, creating a set of rules that ultimately leads to a complete classification system.

Due to its simplicity, speed and the absence of a training phase, the K-nearest neighbor (KNN) algorithm is one of the most used techniques in machine learning. Based on a neighborhood majority voting scheme, the classification of a given sample is assigned to the most common class amongst its K-nearest neighbors. Interesting results have been shown from its use by Pirttikangas et al. (2006) and Preece et al. (2009a).

Another frequently used approach is based on the Bayes' rule. Regarding the simplicity, the Naive Bayes (NB) algorithm may be a suitable approach as long as the stochastical independence is guaranteed, which in practice is normally attained. Ravi et al. (2005), Lester et al. (2005), and Maurer et al. (2006) have used it in different activity-recognition-related problems.

Support vector machines (SVM) is a machine-learning technique which has become very popular in the last years. The promising results recently obtained in previous studies as Parera et al. (2009) or He and Jin (2009) reinforce the idea of its use.

3.2 Fusion

There are different reasons supporting the use of meta-classifiers or decision fusion models in the activity-recognition field. One of the most important is to increase the efficiency and accuracy of the recognition system. A multistage combination of rules in principle allows us to fuse the decision of simple classifiers using a small set of cheap features. Unlike signal or feature-level fusion approaches, the fusion at the classification level reduces the required complexity at the lower levels of the ARC, increasing as well the robustness and adaptability of the recognition system (Zappi et al. 2007). Moreover the models may be, in general, scaled to a large number of sensors, even supporting the combination of heterogeneous sensors (Ward et al. 2006).

Even whether the fusion may be just defined at the sensor level it may also be interesting to define it in terms of classes (here activities). It has been demonstrated that binary or class-specialized classifiers, in general, perform better than direct multiclass entities (Allwein et al. 2000). The idea then is to combine the decisions of simple binary classifiers, specialized in the insertion or rejection of a particular class. This reduces, in general, the feature-level requirements through the use of more classification entities,

which nevertheless are completely defined through a few parameters once trained.

For the decision fusion, different schemes have been proposed in the literature (Kittler et al. 1998). In this paper, we focus on two fusion techniques which have been previously used in the activity-recognition field. The third one is originally proposed as an alternative to the former ones.

3.2.1 Hierarchical decision (HD)

Some classification entities may work better on the recognition of some particular classes. Consequently it may be reasonable to rank the classifiers' decisions (o the classification entities' decisions). The idea is to give more importance to those classifiers which generally behave better, so allowing them to decide first. Thus, the decisions are made in strict order of classification capabilities (the ranking is established according to performance criteria). If the decision is negative, the next classifier in the hierarchy is asked and so on.

The main problem of this model is the dependence of the decisions of the low-level entities on the upper ones. If an upper-level decision-maker fails, the error propagates down the hierarchy and the final decision is likely erroneous. It is true that the decision error is minimized with the hierarchical configuration, since the best decision entities are on top of the hierarchy. However, the potential decisions of the upper-level entities basically determine the eventually adopted decision, even when the rest of the entities may be only slightly less accurate. This translates into a model behaving satisfactorily when at least a few entities are reliable enough, but may neglect the potential of the rest of the lower-ranked decision entities.

3.2.2 Majority voting (MV)

To give the same opportunities to all the decision entities, a democracy-based model may be defined. The majority voting or plurality voting is a naive approach relying on an equality scheme. The final adopted decision is the one obtaining more votes from the participant decision entities. The main properties of this method are the fairness and decisiveness, which translate in a similar treatment of each vote and eventually a unique resulting decision.

When considering rich sensor environments, the use of plural decision may be particularly recommended. Nevertheless, the main risk of this approach is precisely related to the previous highlighted properties. The same significance is given to all the decisions even when they may be different in accuracy. Consequently, a high degradation on the performance is expectable in noisy environments ("tyranny of the majority"): a minority of high-accuracy decision-makers can be hidden by a majority of weak-decision makers. Such kind of situations are typically depicted in context and activity recognition, since the sensors may fail or the setups change, with a more remarkable effect when the dimensionality of the sensor topology is reduced (Zappi et al. 2007).

3.2.3 Hierarchical-weighted classification (HWC)

The model we propose is a fusion technique which takes in the advantages of the hierarchical decision and majority voting models. The idea is to give to all the entities the opportunity of collaborating on the decision making, but ranking the relative importance of each decision through the use of weights based on the individual performance of each entity.

The HWC is composed by three classification levels or stages (see Fig. 2). In general, for M sources of information (sensors) and N-classes (activities), a set of M by N "class classifiers" $(c_{mn}, \forall m = 1, \ldots, M, n = 1, \ldots, N)$ is defined. They are binary classifiers specialized in the insertion/rejection of the class n using the data obtained from the mth source. Each one applies a one-versus-rest strategy[1], and any type of classification paradigm may be used. This defines the base-level or class-level classifier. The second stage, source-level classification, is here defined by M "source classifiers" $(S_m, \forall m = 1, \ldots, M)$. Source classifiers are not machine-learning type classifiers, but hierarchical decision models defining a classification entity strictly speaking. These structures are composed by several class classifiers as shown in Fig. 2, whose decisions are combined through a weighting scheme. The model is replicated at the next level, ultimately defining a decision structure constituted by the fusion of source classifiers.

In accordance to the structure described above, a process consisting of a few main steps is carried out to build the complete HWC. The process starts by evaluating the individual average accuracy of each class classifier $(\overline{R_{mn}})$. A p fold cross-validation is suggested to accomplish this task. The whole process is repeated for each source, and a weight is then obtained for each class classifier:

$$\beta_{mn} = \frac{\overline{R_{mn}}}{\sum_{k=1}^{N} \overline{R_{mk}}} \tag{1}$$

These weights represent the importance that each class classifier will have on the source classifier decision scheme. A specific voting algorithm is considered at this stage to fuse the class classifier decisions into a single decision for each source, respectively. For a source m, given a sample x_{mk} to be classified and being q the class predicted by the

---

[1] Other approaches as the one-versus-one may be similarly applied, but here, the one-versus-rest is particularly recommended to reduce the number of classification entities.
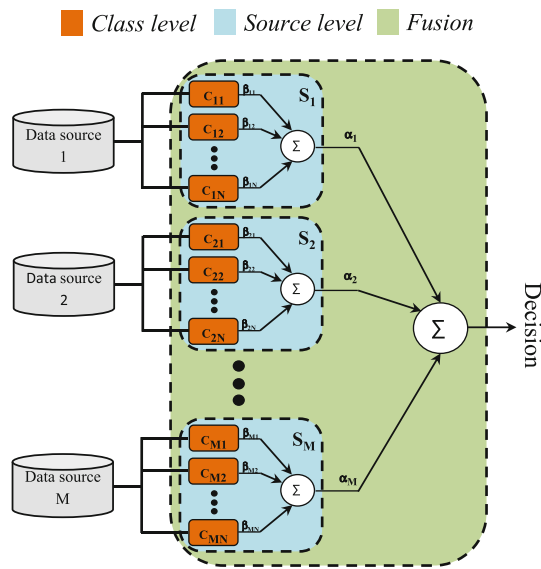
Fig. 2 Structure of the hierarchical weighted classifier. Problem with N-classes and M sources

classifier c_{mn}, if such class belongs to the class of specialization ($q = n$), the classifier will set its decision to '1' for the class n and '0' for the rest of the classes. The opposite is made for ($q \neq n$). The decision of the classifier n for the class q is given by ($\forall \{q, n\} = 1, \ldots, N$):

$$D_{nq}(x_{mk}) = \begin{cases} 1, x_{mk} \text{ classified as } q & \\ 0, x_{mk} \text{ not classified as } q & (\forall q = n) \\ 1, x_{mk} \text{ not classified as } q & \\ 0, x_{mk} \text{ classified as } q & (\forall q \neq n) \end{cases} \quad (2)$$

Now the output of the mth source classifier may be computed as follows:

$$O_{mq}(x_{mk}) = \sum_{n=1}^{N} \beta_{mn} D_{nq}(x_{mk}) \quad (3)$$

The class predicted for the mth source classifier (q_m) is the class q for which the source classifier output is maximized:

$$q_m = \underset{q}{\mathrm{argmax}}(O_{mq}(x_{mk})) \quad (4)$$

At this stage, the source-level classifiers are completely defined. Every source classifier may be independently used. However, as introduced in Sect. 3.2, the fusion of sources has been devised to be, in general, a more robust and efficient solution. Furthermore it provides the capability of being used in different scenarios without needing a retraining. Similarly to Eq. 1, a weight is now calculated for each source by first assessing the average accuracy rates of each source classifier ($\overline{R_m}$). This is performed through a new cross-validation process for the already trained source classifiers. The weight for each source is:

$$\alpha_m = \frac{\overline{R_m}}{\sum_{k=1}^{M} \overline{R_k}} \quad (5)$$

The output is calculated taking into account the individual outputs obtained from each source classifier. For a sample x_k defined through the corresponding samples obtained from each source ($x_{1k}, \ldots, x_{Mk}$), the output is:

$$O_q(x_k) = O_q(\{x_{1k}, \ldots, x_{Mk}\}) = \sum_{p=1}^{M} \alpha_p O_{pq}(x_{pk}) \quad (6)$$

Similar to (4) the eventually assigned class q is obtained as:

$$q = \underset{q}{\mathrm{argmax}}(O_q(x_k)) \quad (7)$$

At this point the HWC is simply defined through the trained class classifiers (c_{mn}), the class-level weights (β_{mn}) and the source-level weights (α_m).

4 Results and analysis

4.1 Experimental setup

To compare the capabilities of the described models, a well-characterized inertial sensor-based dataset, highly cited in the activity-recognition field, is used (Bao and Intille 2004). It comprises the acceleration data registered for 20 subjects aged 17–48 while performing a set of daily living activities. From the whole set, the most representative nine are selected (Fig. 3), covering from intense activities such as running or cycling to fitness exercises like stretching, or relaxed activities such as sitting or lying down. The movements were recorded through five biaxial accelerometers attached to the subjects' right hip, dominant wrist, non-dominant arm, dominant ankle and non-dominant thigh, respectively.

4.2 Methods

According to the ARC presented in Fig. 1, the data are first preprocessed. The recorded signals are affected by spurious spikes, offset discontinuities and electronic noise. A 20 Hz cutoff low-pass elliptic FIR filter is used to remove such anomalies. This is supported by Bouten et al. (1997) and Mathie et al. (2004) who state a 20 Hz sampling rate is sufficient to assess habitual daily physical activity. The signals are subsequently partitioned in windows of data of approximately 6 s as suggested in Bao and Intille (2004). It is assumed an independent system provides the start and end point of the actions.

To analyze the required classification complexity for each presented method, different feature vector lengths are
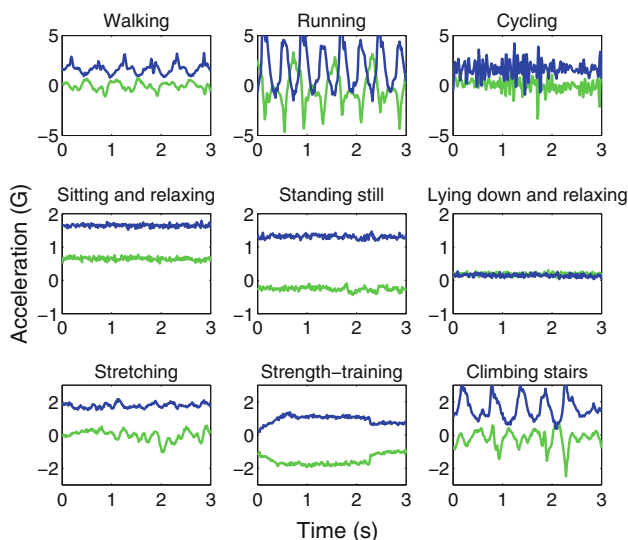
Fig. 3 Selected examples for the nine recorded activities. The data correspond to the acceleration signals (*green*, *X* axis and *blue*, *Y* axis) registered through the arm sensor for one particular subject (color figure online)

tested (1, 5, 10 and 20 features). Here, a subset of the complete set of features proposed in a previous work is considered (Banos et al. 2012). For that purpose, the best features ranked for each sensor are accordingly selected through the use of a ROC feature selector (Theodoridis and Koutroumbas 2008), until reaching the feature vector lengths defined for each case.

The classification process is carried out through the use of the different methods presented in Sect. 3. The machine-learning paradigms used for both direct and fusion models correspond to those presented in Sect. 3.1. In particular, a C4.5 implementation (Duda et al. 2000) is used for the DT. For the SVM an RBF kernel with automatically tuned (grid search) hyper-parameters γ and C is used (Cristianini and Shawe-Taylor 2000). Likewise the K values for the KNN models (Cover and Hart 1967) are obtained. Finally, the NB approach presented in Theodoridis and Koutroumbas (2008) is considered.

Binary or class-specialized classifiers are trained for the fusion approaches, while multiclass models are used for the direct approaches. In all cases, a tenfold cross-validation process is applied. A reserved subset of the training data (normally representing a 30 % of the initial dedicated data) is particularly used for the HWC weights assessment. The process is repeated 100 times for each method to ensure statistical robustness.

4.3 Evaluation

Figure 4 depicts the accuracy differences among the direct multiclass models, based on the most accurate sensor (here

the wrist), and those based on the fusion of all the available sensors (the proposed HWC technique is considered for that purpose). Clearly the fusion approach systematically outperforms the direct ones with independence of the considered machine learning paradigm. This demonstrates the potential of the fusion of different sensors with respect to the use of just one sensor. Moreover, we want to stress that more than 95 % accuracy is obtained for some fusion models based on KNN or SVM through the use of one single feature for each class-level classifier. This implies an outperformance of up to 20 % for the KNN-based models. Even whether an increment in the number of features allows, in general, the models to improve their recognition capabilities, up to 20 features are required for the best direct model to achieve an accuracy comparable to the one achievable by means of the fusion.

In Fig. 5, the performance of the different fusion models is presented. The confusion matrices for each paradigm are shown, since these are more informative than the accuracy rates (in any case calculable from the "diagonalness" of the matrices). A general overview allows us to confirm the HWC model as the most reliable. Even when a single feature is used for each classification entity the confusion matrices are almost diagonal, thereby demonstrating a high accuracy level. An increment on the number of features improves, in general, the classification capabilities but for the MV. The observed enhancement of the HD models may be explained with the results previously analyzed for Fig. 4. Since better individual sensor recognition systems are generally obtained as more features are used, the hierarchy down-propagated error is minimized. This determines the decisions are, in general, made on top of the hierarchy, and the error reduced. From a statistical point of view, MV is easily corrupted when the number of low-performance decision entities overtake the accurate ones.
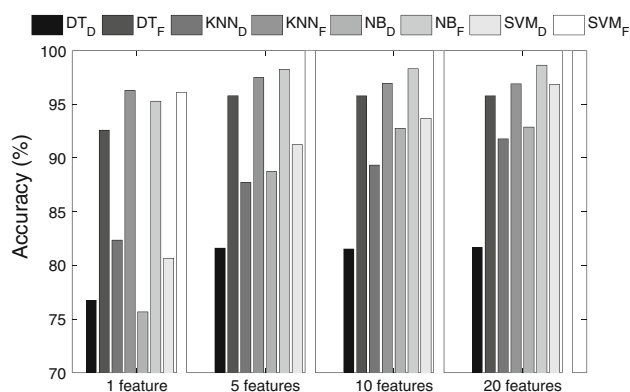

Fig. 4 Comparison between the direct multiclass (*D*) approach and the HWC fusion (*F*) approach. Legend: <classification paradigm> <decision model>. The wrist-based sensor is used for the direct approach since it renders the best performance of all tested sensors

In such circumstances, the potential of those entities offering a high performance may be hidden by a majority of less-accurate classification entities, introducing a systematic error which degrades the performance of the whole recognition system.

With regard to the discriminant capabilities at the activity level, different results are also depicted. Particularly, diverse are the misclassifications for the HD model (more as the feature vector length reduces). KNN with one feature and SVM with five or less provide the poorest results, but the best when ten or more features are considered. A reduced feature vector suffices for some activities as "sitting and relaxing", but it is not enough for most of the rest. In such case, insertions and rejections are mainly determined by a few high-ranked classifiers, thereby hiding the decisions provided by a majority of less accurate class-specialized classifiers. This is encountered to coincide with the classes with more misclassifications. For MV the decisions seem to be biased, systematically misclassified as "walking". This is not because there exists a

Fig. 5 Confusion matrices for each fusion modality and machine learning paradigm. Different feature vector lengths are considered. *Legends*: *1* walking, *2* running, *3* cycling, *4* sitting and relaxing, *5* standing still, *6* lying down and relaxing, *7* stretching, *8* strength-training, *9* climbing stairs

tendency towards interpreting the considered activities as such class, but rather due to the source-level class criterion selection. When all class classifiers reject the class, all classes receive the same number of votes, thereby on equal terms, the first class in order (i.e., "walking") is the eventually selected. Fortunately both situations are efficiently overcome when the HWC is considered. The weighting of the decisions provided at the activity level is definitely important to avoid the misclassifications in the above cases. Indeed, the performance is almost absolute but for "stretching" and "strength training" activities, which are a little less distinguishable. This is probably due to the motion similarities between these two activities.

One of the major drawbacks of the fusion approaches is the drop on the performance for small sensor networks. To prove the scalability of the proposed technique, Fig. 6 depicts the confusion matrices when different number of sensors is used. We show all the possible combinations of sensors, using the HWC with SVM and a single feature for the class-level entities (the simplest realization). Moreover, this analysis also helps us to identify which body sensor locations are the most interesting for the recognition task.
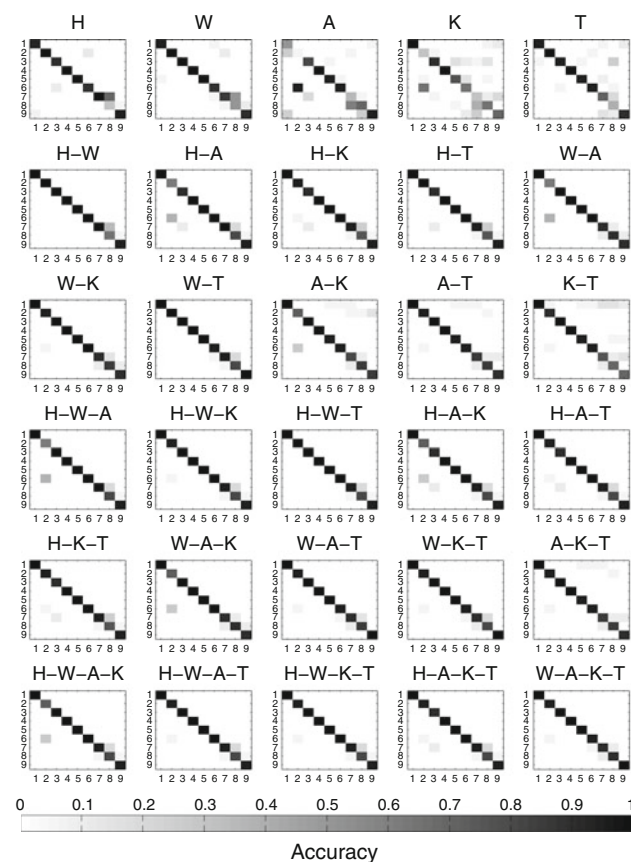
Fig. 6 Confusion matrices for the HWC fusion model when all the possible sensor combinations are considered. SVM is used as machine learning model. Legend of the sensors: *H* hip, *w* wrist, *A* arm, *K* ankle, *T* thigh

The wrist sensor is shown to be the most reliable of the considered five. The fusion at the class-level works reasonably well but for the arm (A) and ankle (K) sensors. Conversely to the other previously analyzed fusion approaches, the combination of good-quality sensors with poor-performance sensors generally results in a better recognition system. Furthermore, the combination of sensors which in principle do not provide high levels of individual performance translates into a high discriminant model (e.g., combination A–K). This is particularly interesting for those cases when the sensors' performance degrades due to unexpected variations in the devices or their deployment. At any rate, the performance is clearly improved as more sensors are considered, with almost absolute accuracy for combinations of three sensors such as the hip, wrist and thigh.

5 Discussion

The comparison between HWC and the rest of models here presented exhibits the remarkable capabilities of the proposed fusion technique. With regard to the accuracy of the systems, it has been ascertained that the decisions provided by the collectivity are more reliable than the one dependent on one single source. That is completely reasonable, since some of the activities may higher involve a subset of the considered sensors as each part of the body moves differently. The natural alternative is to define the fusion at the first levels of the ARC (signal level or feature level), but systems based on such approach are highly constrained, since a change on the sensor topology require a retraining of the whole system (which in principle is not affordable in an online context). Conversely, when the fusion is performed on top of the ARC no influence on other stages is experienced. The addition or removal of sensors and classes basically translates into the inclusion or elimination of the associated knowledge-inference entities, but the rest of the originals remain identical. Then to include these changes, the fusion parameters are just updated. This may be performed in a negligible time, thereby allowing for the system updating even in an online manner. These relevant characteristics define HWC structurally simple and flexible.

Now, with respect to other fusion approaches which may share the aforementioned features, the obtained results demonstrates HWC stands out above the rest. MV and HD have been shown not to perform well due to different causes. The potential of the MV approach is restricted by a majority of less-accurate recognition entities, while HD requires at least one decision entity performing with a significant accuracy level (normally obtained when a substantial number of features are considered). The HWC

model has been demonstrated to be both scalable and efficient, since the model outperforms the accuracy of the individual recognition entities independently of the number of sensors considered. Furthermore, this is not just restricted to the sensor level but extended to the class level. In a previous study, we showed a first attempt for a four activities-based recognition problem (Banos et al. 2011). In this work, we have likewise extended the recognition capabilities of the system to a complete set of nine activities, some of them likely similar from a body-motion point of view. These outstanding results also reinforce the use of the weighting at the activity or class level.

6 Conclusion

We have presented a fusion technique which combines the capabilities of simple classification entities at the class (activity) and source (sensor) levels. The model outperforms direct multiclass approaches with a considerable reduction in the dimensionality of the required feature vector.

The use of a hierarchical weighting decision scheme has been demonstrated to significantly improve the scalability and robustness with respect to other traditional fusion techniques. The combination of poor decision entities leads to a decision system which, in general, behaves better or at worst as the best of the constituent entities.

The benefits of the presented methodology could be similarly envisioned in changeable scenarios. In principle, with the fusion at the classification level, the addition or removal of classes and/or sensors do not imply to modify and/or retrain the whole system. Only the corresponding entities should be added or removed with a subsequent updating of the fusion parameters, which might be performed at runtime and without disrupting the normal use of the recognition system.

References

Allwein EL, Schapire RE, Singer Y, Kaelblin P (2000) Reducing multiclass to binary: a unifying approach for margin classifiers. J Mach Learn Res 1:113–141

Amft O, Lukowicz P (2009) From backpacks to smartphones: past, present, and future of wearable computers. Pervasive Computi IEEE 8(3):8–13

Banos O, Damas M, Pomares H, Rojas I (2011) Automatic recognition of daily living activities based on a hierarchical classifier. In: Advances in computational intelligence: 11th international work-conference on artificial neural networks, vol 6692. Springer, Berlin, pp 185–193

Banos O, Damas M, Pomares H, Prieto A, Rojas I (2012) Daily living activity recognition based on statistical feature quality group selection. Expert Syst Appl 39(9):8013–8021

Bao L, Intille SS (2004) Activity recognition from user-annotated acceleration data. Pervasive Comput 23:1–17

Bonomi AG, Plasqui G, Goris AHC, Westerterp KR (2009) Improving assessment of daily energy expenditure by identifying types of physical activity with a single accelerometer. J Appl Physiol 107(3):655–661

Bouten C, Koekkoek K, Verduin M, Kodde R, Janssen J (1997) A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. Biomed Eng IEEE Trans 44(3):136–147

Cover T, Hart P (1967) Nearest neighbor pattern classification. Inf Theory IEEE Trans 13(1):21–27

Cristianini N, Shawe-Taylor J (2000) An introduction to support vector machines: and other kernel-based learning methods. Cambridge University Press, New York

Duda RO, Hart PE, Stork DG (2000) Pattern classification, 2nd edn. Wiley-Interscience, New York

Ermes M, Parkka J, Mantyjarvi J, Korhonen I (2008) Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. IEEE Trans Inf Technol Biomed 12(1):20–26

European-Commission (2011) Ambient assisted living joint programme. http://www.aal-europe.eu/

He Z, Jin L (2009) Activity recognition from acceleration data based on discrete consine transform and svm. In: Systems, man and cybernetics, 2009. SMC 2009, IEEE international conference, pp 5041–5044

Kittler J, Hatef M, Duin R, Matas J (1998) On combining classifiers. IEEE Trans Pattern Anal Mach Intell 20(3):226–239

Lester J, Choudhury T, Kern N, Borriello G, Hannaford B (2005) A hybrid discriminative/generative approach for modeling human activities. In: Proceedings of the 19th international joint conference on artificial intelligence, IJCAI'05. Morgan Kaufmann Publishers Inc., San Francisco, pp 766–772

Mantyjarvi J, Himberg J, Seppanen T (2001) Recognizing human motion with multiple acceleration sensors. In: Systems, man, and cybernetics, 2001 IEEE international conference, vol 2, pp 747–752

Mathie MJ, Coster ACF, Lovell NH, Celler BG (2004) Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement. Physiol Meas 25(2):1–20

Maurer U, Smailagic A, Siewiorek D, Deisher M (2006) Activity recognition and monitoring using multiple sensors on different body positions. In: Wearable and implantable body sensor networks, 2006. BSN 2006. International workshop on, pp 113–116

Parera J, Angulo C, Rodriguez-Molinero A, Cabestany J (2009) User daily activity classification from accelerometry using feature selection and svm. LNCS, vol 5517, pp 1137–1144

Parkka J, Ermes M, Korpipaa P, Mantyjarvi J, Peltola J, Korhonen I (2006) Activity classification using realistic data from wearable sensors. IEEE Trans Inf Technol Biomed 10(1):119–128

Pirttikangas S, Fujinami K, Seppanen T (2006) Feature selection and activity recognition from wearable sensors. In: Ubiquitous computing systems, third international symposium, LNCS, vol 4239, pp 516–527

Preece S, Goulermas J, Kenney L, Howard D (2009) A comparison of feature extraction methods for the classification of dynamic

activities from accelerometer data. IEEE Trans Biomed Eng 56(3):871–879

Preece SJ, Goulermas JY, Kenney LPJ, Howard D, Meijer K, Crompton R (2009b) Activity identification using body-mounted sensors-a review of classification techniques. Physiol Meas 30(4):1–33

Ravi N, Mysore P, Littman ML (2005) Activity recognition from accelerometer data. In: In Proceedings of the seventeenth conference on innovative applications of artificial intelligence, pp 1541–1546

Roggen D, Calatroni A, Rossi M, Holleczek T, Forster K, Troster G, Lukowicz P, Bannach D, Pirkl G, Ferscha A, Doppler J, Holzmann C, Kurz M, Holl G, Chavarriaga R, Sagha H, Bayati H, Creatura M, del Millan JR (2010) Collecting complex activity datasets in highly rich networked sensor environments. In: Networked sensing systems (INSS), 2010 seventh international conference, pp 233–240

Sazonov E, Makeyev O, Schuckers S, Lopez-Meyer P, Melanson E, Neuman M (2010) Automatic detection of swallowing events by acoustical means for applications of monitoring of ingestive behavior. IEEE Trans Biomed Eng 57(3):626–633

Sazonov E, Metcalfe K, Lopez-Meyer P, Tiffany S (2011) Rf hand gesture sensor for monitoring of cigarette smoking. In: Sensing technology, 2011 fifth international conference, pp 426–430

Sharma R, Pavlovic V, Huang T (1998) Toward multimodal human-computer interface. In: Proceedings of the IEEE, vol 86, pp 853–869

Shephard RJ (2003) Limits to the measurement of habitual physical activity by questionnaires. Br J Sports Med 37(3):197–206

Staudenmayer J, Pober D, Crouter S, Bassett D, Freedson P (2009) An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer. J Appl Physiol 107(4):1300–1307

Theodoridis S, Koutroumbas K (2008) Pattern recognition, 4th edn. Academic Press, London

Ward J, Lukowicz P, Troster G, Starner T (2006) Activity recognition of assembly tasks using body-worn microphones and accelerometers. IEEE Trans Pattern Anal Mach Intell 28(10):1553–1567

Warren JM, Ekelund U, Besson H, Mezzani A, Geladas N, Vanhees L (2010) Assessment of physical activity a review of methodologies with reference to epidemiological research: a report of the exercise physiology section of the european association of cardiovascular prevention and rehabilitation. Eur J Cardiovas Prev Rehabil 17(2):127–139

WHO (2005) Global report, preventing chronic diseases: a vital investment tech rep. World Health Organization, Geneva

WHO (2006) American report, regional strategy on an integrated approach to the prevention and control of chronic diseases including diet, physical activity, and health: tech rep. World Health Organization, Geneva

Zappi P, Stiefmeier T, Farella E, Roggen D, Benini L, Troster G (2007) Activity recognition from on-body sensors by classifier fusion: sensor scalability and robustness. In: 3rd international conference on intelligent sensors, sensor networks and information, pp 281–286