Multiwindow Fusion for Wearable Activity Recognition

Oresti Banos[1]*, Juan-Manuel Galvez[2], Miguel Damas[2], Alberto Guillen[2], Luis-Javier Herrera[2], Hector Pomares[2], Ignacio Rojas[2], Claudia Villalonga[2], Choong Seon Hong[1], and Sungyoung Lee[1]

[1] Department of Computer Engineering, Kyung Hee University, Yongin-si, Korea
{*oresti,cshong,sylee*}*@khu.ac.kr*

[2] Department of Computer Architecture and Computer Technology, University of Granada, Granada, Spain
{*jonas,mdamas,aguillen,jherrera,hector,irojas,cvillalonga*}*@ugr.es*

Abstract. The recognition of human activity has been extensively investigated in the last decades. Typically, wearable sensors are used to register body motion signals that are analyzed by following a set of signal processing and machine learning steps to recognize the activity performed by the user. One of the most important steps refers to the signal segmentation, which is mainly performed through windowing approaches. In fact, it has been proved that the choice of window size directly conditions the performance of the recognition system. Thus, instead of limiting to a specific window configuration, this work proposes the use of multiple recognition systems operating on multiple window sizes. The suggested model employs a weighted decision fusion mechanism to fairly leverage the potential yielded by each recognition system based on the target activity set. This novel technique is benchmarked on a well-known activity recognition dataset. The obtained results show a significant improvement in terms of performance with respect to common systems operating on a single window size.

Keywords: Activity recognition, Segmentation, Windowing, Wearable sensors, Ensemble methods, Data fusion

1 Introduction

The identification of human behavior based on body-worn sensors, also known as wearable activity recognition, has attracted very much attention during the last years. Wearable activity recognition systems have been proven of particular interest, for example, to promote healthier lifestyles [1, 26, 2], detect anomalous behaviors [23, 20] or track on conditions [16]. A set of steps combining signal processing and machine learning techniques are normally used in the activity recognition process. Concretely, one or various sensors are typically placed on limbs and trunk to register and translate human body motion into digital signals

* Corresponding author.

representing the magnitude measured, normally acceleration. The registered signals are sometimes filtered when these are found to be disturbed by electronic noise or other type of artifacts [17]. To capture the dynamics of the movement, the signals are subsequently partitioned in segments or data windows [8]. Then a feature extraction process is performed on each data window to provide a handler representation of the signals for the pattern recognition stage. Diverse heuristics [18], time-frequency domain [19, 22] and other sophisticated mathematical and statistical functions [4] are commonly used to that end. In some cases, a feature selector is used to reduce redundancy among features as well as to minimize dimensionality [21]. The resulting feature vector is provided as input of a classifier, which ultimately yields the recognized activity or class to one of the considered for the particular application. All these steps, commonly referred as to activity recognition chain, are extensively reviewed in [10].

Although all stages of the activity recognition process are undoubtedly important, a recent work [8] showed the particular impact of the segmentation phase on the accuracy of the recognition models. Amongst other findings, this work showed the existing relation among activity categories and involved body parts with the window size utilized during the segmentation process. As a result, specific design figures are proposed, which in principle allow developers to set a certain window size value to optimize the recognition system capabilities. Nevertheless, these values are very specific to each application domain; thus, no particular window size may exist for systems intended to recognize multiple diverse activities. In that vein, this paper proposes the use of fusion mechanisms to benefit from the utilization of several window sizes instead of restricting to a single one. Fusion strategies have been already used in previous activity recognition systems for diverse purposes, such as dealing with sensor displacement [6, 9], anomalies [3] and power management [27]. This work presents an innovative multiwindow fusion technique that weights and combines the decisions provided by multiple activity recognizers configured to operate on different windows sizes of the same input data. The rest of the paper is organized as follows. Section 2 describes the multiwindow fusion method. Section 3 presents and discusses the results obtained after benchmarking the proposed method on a well-known activity recognition dataset. Final conclusions are summarized in Section 4.

2 Multiwindow fusion

As stated before, the recognition of activities of diverse characteristics potentially require the use of various levels of segmentation. Therefore, the model proposed here consists in the combination of multiple activity recognition chains, every one utilizing a different window size configuration. Each of these chains builds on the same input signals, and for the sake of simplicity, all are considered to use similar feature extraction and classification procedures. For practical reasons, the selected window sizes should be divisors of the largest one among considered, which is defined according to the particular needs posed by the target activity set and system recognition period. The key challenge of this approach consists in the intelligent aggregation of the decisions, i.e., recognized activities, delivered by each chain. To that end, a two-step fusion process is here suggested. First, the decisions provided by each individual activity recognizer are locally weighted and aggregated to yield a sole recognized activity per chain. The activities identified for each chain are then combined in a second stage to eventually deliver a unique recognized activity. The complete structure of the proposed model is depicted in Figure 1, while its mathematical foundation is described in the following.

Let us consider a problem with N classes or activities, $n = 1, ..., N$. Given a set of raw, u, or preprocessed, p, sensor data, these are segmented by using Q different window sizes, $\{W_1, ..., W_{Q-1}, W_Q\}$, with $W_Q|W_{Q-1}|...|W_1$ divisors of W_Q, and W_Q formally representing the system recognition period. This leads to the creation of Q independent recognition chains, in which every data window of size W_Q, i.e., s^{W_Q}, is split into W_Q/W_k segments of size W_k, i.e., $\{s_1^{W_k}, ..., s_i^{W_k}, ..., s_{W_Q/W_k}^{W_k}\}$, for all $k = 1, ..., Q$ and $i = 1, ..., W_Q/W_k$. Each segment $s_i^{W_k}$ is transformed into features, $f(s_i^{W_k})$, which are input to each respective classifier, yielding a recognized activity or class, $c_i^{W_k}$.

At this point the multiwindow fusion technique is employed. First, the decisions of each individual classifier are weighted and averaged across all segments and for all classes:

$$WD_n^{W_k} = \sum_{i=1}^{W_Q/W_k} \lambda_n^{W_k} \qquad \forall\, c_i^{W_k} = n \tag{1}$$

where the weight factor $\lambda_n^{W_k}$ represents the capabilities of the classifier k, that operates on data windows of size W_k, for the recognition of the activity or class n. This factor is different for each class and window size, and can be calculated from a prior evaluation of the performance of each respective activity recognition chain, similarly as it is proposed in [5]. Now, the class c^{W_k} predicted for each classifier after fusion is the class n for which $WD_n^{W_k}$ is maximized:

$$c^{W_k} = \underset{n}{\operatorname{argmax}} \left(WD_n^{W_k} \right) \tag{2}$$

This process is repeated in a second level by weighting and averaging the decisions obtained in the previous fusion step for each respective window size:

$$WD_n = \sum_{k=1}^{Q} \lambda_n^{W_k} \qquad \forall\, c^{W_k} = n \tag{3}$$

The eventual recognized class is defined as the one obtaining the highest weighted sum:

$$c = \underset{n}{\operatorname{argmax}} \left(WD_n \right) \tag{4}$$

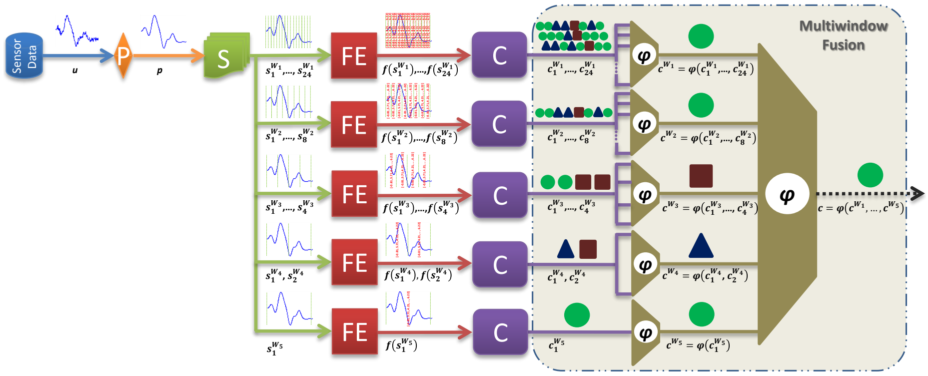

Fig. 1. Multiwindow fusion schema (example for Q=5 different window sizes). The raw *sensor data*, u, is *preprocessed*, p, and *segmented* into data windows, $s_i^{W_k}$, of size W_k, thus defining Q independent recognition chains. For each chain k and window i, a set of *features* are *extracted*, $f(s_i^{W_k})$, which are input to each respective classifier, yielding a recognized activity or class, $c_i^{W_k}$. The recognized classes are weighted and fused in a first stage to predict the most likely activity for each chain, $c_i^{W_k}$. The classes predicted after fusion for each chain are again weighted and fused to deliver the eventual recognized activity.

3 Results and discussion

3.1 Experimental setup

One of the most complete available activity recognition datasets [7] is used for evaluation. This dataset comprises motion data, namely, acceleration, rate of turn and magnetic field orientation, recorded for 17 volunteers while performing 33 fitness activities. A set of nine inertial sensors attached to different parts of their bodies was used for the motion recording. From all measured magnitudes only the acceleration data is here considered since this demonstrates as the most prevalent sensor modality in previous activity recognition contributions. The potential of this dataset stems from the number of considered activities, diversity of body parts involved, as well as the variety in intensity and dynamicity of the actions. Moreover, all the recordings were collected in an out-of-lab environment with no constraints on the way the activities must be executed.

The activity recognition models devised for evaluation are described next. No preprocessing of the data is applied to avoid the removal of relevant information. This is normal practice when the activities are of a diverse nature. Five window sizes used in previous works are considered for study, respectively, $W_1 = 0.25$, $W_2 = 0.75$, $W_3 = 1.5$, $W_4 = 3$ and $W_5 = 6$, all in seconds. Mean and standard deviation are used for the feature extraction, given their discrimination potential and ease of interpretation in the acceleration domain [13, 14]. Four well-known machine learning techniques widely utilized in previous activity recognition problems are considered for classification, namely, C4.5 decision trees

(DT, [12]), k-nearest neighbors (KNN, [11]), naive Bayes (NB, [25]) and nearest centroid classifier (NCC, [15]). The k-value for the KNN model is particularly set to three as it has been shown to provide good results in related works. The $\lambda_n^{W_k}$ weights used in the fusion process correspond to the *F-score* [24] values obtained in [8] from the analysis of similar recognition systems operating on the window sizes and activities considered in this experiment. The evaluation of the multiwindow fusion models is performed through a ten-fold random-partitioning cross validation process applied across all subjects and activities. The process is repeated 100 times for each method to ensure statistical robustness.

3.2 Multiwindow fusion evaluation

The results obtained for the multiwindow fusion process after assessment of all possible combinations of the selected window sizes are presented in Table 1. No fusion is explicitly performed for the single-window-based recognition models; thus, the results presented for this case refer to the performance obtained at the classification level, i.e., before fusion.

In broad strokes, the use of multiple window sizes certainly improves the recognition capabilities of the considered systems. This result is observed for all classification paradigms. For example, an enhancement of more than 7% is attained when using the combination $W_1W_2W_4$ and DT with respect to the best results obtained by using a single window size, here for W_3. More modest improvements, around 2%, are achieved for NB, NCC, and KNN in similar conditions. The differences are more striking when compared with the worst performing single-window-based recognition models, with improvements of up to 30%.

Another fact to be noted corresponds to the number of windows required for improving the performance of the recognition system. Best results are not necessarily obtained for the combination that involves the highest number of windows. Conversely, in some cases such as for NCC and KNN, the combination of simply two windows turns to be enough to neatly improve the recognition capabilities of the system. This demonstrates the potential of the fusion mechanism even for small sets of decision makers.

As it may be apparent, the use of multiple windows translates into a higher computation complexity, therefore might not be justified under some circumstances or not be recommended when the improvement is negligible. However, in some cases it is observed that the use of multiple windows can actually reduce the recognition time, a key characteristic in applications that require a fast response (e.g., fall detector). This is the case, for example, of the combination W_1W_4 in NB, which enhances the accuracy with respect to the best single-window-based recognition system, W_5, thus permitting to reduce the recognition period from 6s to 3s. The importance of this effect is also observed for the case of W_1W_2 in KNN, which improves the performance of W_4 while reducing the recognition time from 3s to 0.75s.

Finally, it is worth noting that the combination of two or more window sizes generally translates into a recognition performance greater or equal to the one

Combined window sizes	DT	NB	NCC	KNN
W_1	0.835	0.702	0.596	0.976
W_2	0.879	0.868	0.807	0.979
W_3	0.895	0.900	0.864	0.981
W_4	0.886	0.908	0.873	0.984
W_5	0.869	0.910	0.870	0.942
W_1W_2	0.878	0.855	0.760	0.991
W_1W_3	0.915	0.905	0.856	**0.996**
W_1W_4	0.920	0.922	0.870	0.976
W_1W_5	0.915	0.917	0.867	0.967
W_2W_3	0.877	0.905	0.858	0.991
W_2W_4	0.910	0.925	0.878	0.981
W_2W_5	0.917	0.918	0.866	0.958
W_3W_4	0.876	0.925	**0.881**	0.976
W_3W_5	0.893	0.922	0.876	0.954
W_4W_5	0.861	0.923	0.878	0.952
$W_1W_2W_3$	0.954	0.893	0.832	0.995
$W_1W_2W_4$	**0.968**	0.916	0.855	0.990
$W_1W_2W_5$	0.960	0.911	0.855	0.968
$W_1W_3W_4$	0.956	0.927	0.880	0.989
$W_1W_3W_5$	0.956	0.922	0.876	0.967
$W_1W_4W_5$	0.936	0.926	0.878	0.966
$W_2W_3W_4$	0.945	**0.928**	0.880	0.989
$W_2W_3W_5$	0.944	0.923	0.876	0.966
$W_2W_4W_5$	0.928	0.924	0.873	0.966
$W_3W_4W_5$	0.925	0.926	0.878	0.959
$W_1W_2W_3W_4$	0.967	0.926	0.879	0.989
$W_1W_2W_3W_5$	0.961	0.923	0.873	0.967
$W_1W_2W_4W_5$	0.955	0.926	0.874	0.967
$W_1W_3W_4W_5$	0.953	0.927	0.878	0.964
$W_2W_3W_4W_5$	0.945	0.924	0.878	0.962
$W_1W_2W_3W_4W_5$	0.960	0.924	0.878	0.968

Table 1. Multiwindow fusion performance ($F-score$) for all possible combinations of considered window sizes ($W_1 = 0.25s$, $W_2 = 0.75s$, $W_3 = 1.5s$, $W_4 = 3s$ and $W_5 = 6s$) and diverse classification paradigms (DT, NB, NCC, KNN).

of best characteristics among considered. This fact is of special importance since it proves the stability and consistency of the proposed fusion mechanism.

4 Conclusions

The choice of window size used in typical activity recognition applications is highly coupled to the particular characteristics of the activities to be recognized. Previous works proved that a single window size value can be considered in recognition systems devised for a very specific domain including a few similar activities. However, no clear value can be determined for problems involving several activities of a more diverse nature. To overcome this limitation, the simultaneous use of multiple window sizes is here suggested. Concretely, this work proposes a novel multiwindow fusion technique that weights and combines the decisions provided by multiple activity recognizers configured to operate on different windows sizes of the same input data. The proposed approach is shown to significantly outperform classical single-window-based recognition models. Moreover, the performed evaluation also shows that using several windows sizes not necessarily translates into best results, but that considering a few ones might be enough for obtaining a highly accurate recognition system.

Acknowledgments. This work was partially supported by the Industrial Core Technology Development Program, funded by the Korean Ministry of Trade, Industry and Energy, under grant number #10049079. This work was also funded by the Junta de Andalucia Project P12-TIC-2082.

References

1. N. Alshurafa, Wenyao Xu, J.J. Liu, Ming-Chun Huang, B. Mortazavi, C.K. Roberts, and M. Sarrafzadeh. Designing a robust activity recognition framework for health and exergaming using wearable sensors. *IEEE Journal of Biomedical and Health Informatics*, 18(5):1636–1646, 2014.
2. O. Banos, M. Bilal-Amin, W. Ali-Khan, M. Afzel, T. Ali, B.-H. Kang, and S. Lee. Mining minds: an innovative framework for personalized health and wellness support. In *Int. Conf. on Pervasive Computing Technologies for Healthcare*, 2015.
3. O. Banos, M. Damas, A. Guillen, L.-J. Herrera, H. Pomares, I. Rojas, and C. Villalonga. Multi-sensor fusion based on asymmetric decision weighting for robust activity recognition. *Neural Processing Letters*, pages 1–22, 2014.
4. O. Banos, M. Damas, H. Pomares, A. Prieto, and I. Rojas. Daily living activity recognition based on statistical feature quality group selection. *Expert Systems with Applications*, 39(9):8013–8021, 2012.
5. O. Banos, M. Damas, H. Pomares, F. Rojas, B. Delgado-Marquez, and O. Valenzuela. Human activity recognition based on a sensor weighting hierarchical classifier. *Soft Computing*, 17:333–343, 2013.
6. O. Banos, M. Damas, H. Pomares, and I. Rojas. On the use of sensor fusion to reduce the impact of rotational and additive noise in human activity recognition. *Sensors*, 12(6):8039–8054, 2012.
7. O. Banos, M. Damas, H. Pomares, I. Rojas, M. A. Toth, and O. Amft. A benchmark dataset to evaluate sensor displacement in activity recognition. In *Proceedings of the ACM Conference on Ubiquitous Computing*, pages 1026–1035, 2012.
8. O. Banos, J.-M. Galvez, M. Damas, H. Pomares, and I. Rojas. Window size impact in human activity recognition. *Sensors*, 14(4):6474–6499, 2014.
9. O. Banos, M. A. Toth, M. Damas, H. Pomares, and I. Rojas. Dealing with the effects of sensor displacement in wearable activity recognition. *Sensors*, 14(6):9995–10023, 2014.
10. A. Bulling, U. Blanke, and B. Schiele. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Comput. Surv.*, 46(3):33:1–33:33, 2014.

11. T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27, 1967.
12. R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification (2nd Edition)*. Wiley-Interscience, 2000.
13. D. Figo, P. C. Diniz, D. R. Ferreira, and J. M. P. Cardoso. Preprocessing techniques for context recognition from accelerometer data. *Personal and Ubiquitous Computing*, 14(7):645–662, 2010.
14. J. R. Kwapisz, G. M. Weiss, and S. A. Moore. Activity recognition using cell phone accelerometers. *Conference on Knowledge Discovery and Data Mining*, 12(2):74–82, 2011.
15. W. Lam, C.-K. Keung, and C. X. Ling. Learning good prototypes for classification using filtering and abstraction of instances. *Pattern Recognition*, 35(7):1491 – 1506, 2002.
16. A. Laudanski, B. Brouwer, and Q. Li. Activity classification in persons with stroke based on frequency features. *Medical Engineering & Physics*, 37(2):180 – 186, 2015.
17. A. Mannini, S. S. Intille, M. Rosenberger, A. M. Sabatini, and W. Haskell. Activity recognition using a single accelerometer placed at the wrist or ankle. *Medicine and Science in Sports and Exercise*, 45(11):2193–2203, 2013.
18. M. J. Mathie, A. C. F. Coster, N. H. Lovell, and B. G. Celler. Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement. *Physiological Measurement*, 25(2):1–20, 2004.
19. U. Maurer, A. Smailagic, D. P. Siewiorek, and M. Deisher. Activity recognition and monitoring using multiple sensors on different body positions. In *International Workshop on Wearable and Implantable Body Sensor Networks*, pages 113–116, 2006.
20. S. Mazilu, U. Blanke, M. Hardegger, G. Tröster, E. Gazit, and J. M. Hausdorff. Gaitassist: A daily-life support and training system for parkinson's disease patients with freezing of gait. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2531–2540, 2014.
21. S. Pirttikangas, K. Fujinami, and T. Seppanen. Feature selection and activity recognition from wearable sensors. In *International Symposium on Ubiquitous Computing Systems*, volume 4239, pages 516–527, 2006.
22. N. Ravi, P. Mysore, and M. L. Littman. Activity recognition from accelerometer data. In *Proceedings of the Conference on Innovative Applications of Artificial Intelligence*, pages 1541–1546, 2005.
23. A. Sama, C. Perez-Lopez, J. Romagosa, D. Rodriguez-Martin, A. Catala, J. Cabestany, D.A. Perez-Martinez, and A. Rodriguez-Molinero. Dyskinesia and motor state detection in parkinson's disease patients with a single movement sensor. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1194–1197, 2012.
24. M. Sokolova and G. Lapalme. A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4):427 – 437, 2009.
25. S. Theodoridis and K. Koutroumbas. *Pattern Recognition*. Academic Press, 4th edition, 2008.
26. G. M Weiss, J. W. Lockhart, T. T. Pulickal, P. T. McHugh, I. H. Ronan, and J. L. Timko. Actitracker: a smartphone-based activity recognition system for improving health and well-being. *SIGKDD Exploration Newsletter*, 2014.
27. P. Zappi, D. Roggen, E. Farella, G. Tröster, and L. Benini. Network-level power-performance trade-off in wearable activity recognition: A dynamic sensor selection approach. *ACM Trans. Embed. Comput. Syst.*, 11(3):68:1–68:30, 2012.