Introducing a New Benchmarked Dataset for Activity Monitoring

Attila Reiss, Didier Stricker
Department of Augmented Vision
German Research Center of Artificial Intelligence (DFKI)
Kaiserslautern, Germany
{firstname.lastname}@dfki.de

Abstract

This paper addresses the lack of a commonly used, standard dataset and established benchmarking problems for physical activity monitoring. A new dataset — recorded from 18 activities performed by 9 subjects, wearing 3 IMUs and a HR-monitor — is created and made publicly available. Moreover, 4 classification problems are benchmarked on the dataset, using a standard data processing chain and 5 different classifiers. The benchmark shows the difficulty of the classification tasks and exposes new challenges for physical activity monitoring.

1. Introduction

Opposed to most established research fields, there is a lack of a commonly used, standard dataset and established benchmarking problems for physical activity monitoring. This work addresses this issue, thus the main contributions are twofold. On the one hand, a new dataset for physical activity monitoring is created and made publicly available. The following specifications — based on experience made in previous work [2], and the limitations of the few available datasets in this field — are defined for this: a wide range of everyday, household and sport activities should be performed by an adequate number of subjects, wearing a few 3D-IMUs and a physiological sensor. On the other hand, this work also presents an initial benchmarking on various defined tasks, showing the difficulty of common classification problems, and exposing some challenges.

2. Data collection

Three inertial measurement units (IMUs) and a heart rate monitor were used as sensors during the data collection. For the inertial measurements, the Colibri wireless IMUs from Trivisio were used. These relatively lightweight and small IMUs contain 3-axis MEMS sensors (two accelerometers, a gyroscope and a magnetometer), all sampled at $100\,\mathrm{Hz}$. To obtain heart rate information, the BM-CS5SR HR-monitor from BM innovations GmbH was used. The sensors are placed onto 3 different body positions. A chest sensor fixation includes one IMU and the heart rate chest strap. The second IMU is attached over the wrist on the dominant arm, and the third IMU on the dominant side's ankle, both are fixed with sensor straps. A Viliv S5 UMPC (Intel Atom Z520 1.33GHz CPU and 1GB of RAM) was used as data collection companion unit.

In total 9 subjects participated in the data collection, 8 males and 1 female. The subjects were mainly employees or students at our research institute, aged $27.22\ \pm 3.31$ years, and having a BMI of $25.11\ \pm 2.62\ \mathrm{kgm}^{-2}$. One subject was left-handed, all the others were right-handed. All subjects have agreed to the usage of recorded data for scientific purposes. The data collection took place in autumn 2011.

Each of the subjects followed a protocol of 12 activities (lie, sit, stand, walk, run, cycle, Nordic walk, iron, vacuum clean, rope jump, ascend and descend stairs), and optionally performed a few other activities (watch TV, computer work, drive car, fold laundry, clean house, play soccer) as well. Over 10 hours of data were collected altogether from the 18 different activities. The dataset is made publicly available, and can be downloaded from `http://www.pamap.org/demo.html`. A brief description of each of the performed activities, and a summary of how much data was recorded per activity, can be also found attached to the published dataset.

3. Benchmarking: methods and results

Standard methods are used for creating the benchmark, the data processing follows a classical approach similar e.g. to the activity recognition chain presented in [3]. The timestamped raw sensory data from the 3 IMUs and the HR-monitor is synchronized in the preprocessing step. This data is segmented using a sliding window of 5.12 seconds with a

Table 1. Performance measures I

	Intensity estimation task				Basic activity recognition task			
	Standard CV		LOSO CV		Standard CV		LOSO CV	
Classifier	F-measure	Accuracy	F-measure	Accuracy	F-measure	Accuracy	F-measure	Accuracy
Decision tree (C4.5)	0.9789	0.9823	0.9426	0.9526	0.9968	0.9970	0.9401	0.9447
Boosted C4.5	0.9986	0.9988	0.9518	0.9587	0.9995	0.9995	0.9794	0.9785
Bagging C4.5	0.9831	0.9866	0.9480	0.9552	0.9970	0.9971	0.9392	0.9433
Naive Bayes	0.8845	0.9310	0.8750	0.9251	0.9921	0.9923	0.9703	0.9705
kNN	0.9986	0.9982	0.9604	0.9666	1.0000	1.0000	0.9938	0.9932

Table 2. Performance measures II

	Background activity recognition task				All activity recognition task			
	Standard CV		LOSO CV		Standard CV		LOSO CV	
Classifier	F-measure	Accuracy	F-measure	Accuracy	F-measure	Accuracy	F-measure	Accuracy
Decision tree (C4.5)	0.9743	0.9709	0.8768	0.8722	0.9558	0.9546	0.8300	0.8244
Boosted C4.5	0.9985	0.9980	0.9433	0.9377	0.9974	0.9969	0.8928	0.8796
Bagging C4.5	0.9823	0.9787	0.9047	0.9042	0.9667	0.9666	0.8556	0.8554
Naive Bayes	0.9105	0.8508	0.8874	0.8308	0.9469	0.9438	0.8362	0.8365
kNN	0.9974	0.9957	0.9443	0.9264	0.9942	0.9925	0.9110	0.8924

shifting of 1 second. From the segmented 3D-acceleration data, various signal features were calculated in both time and frequency domain (mean, variance, energy, etc.), and (normalized) mean and gradient are calculated on the heart rate data [2]. The extracted features serve as input for the next processing step, the classification. Five different classifiers were selected from the Weka toolkit [1] for creating the benchmark: Decision tree (C4.5), Boosted C4.5 decision tree, Bagging C4.5 decision tree, Naïve Bayes and kNN.

Four different classification problems are defined on the 12 protocol-activities. The 'intensity estimation task' defines 3 classes based on the metabolic equivalent: activities of light, moderate and vigorous effort [2]. The 'basic activity recognition task' has 5 activity classes: lie, sit/stand, walk, run and cycle. The 'background activity recognition task' has an additional *other* class containing the remaining 6 activities. Finally, the 'all activity recognition task' defines a separate class for each of the 12 activities.

Table 1 and Table 2 show the performance measures of all 5 classifiers applied to all 4 classification problems. Both subject dependent (standard 9-fold cross-validation) and subject independent (leave-one-subject-out 9-fold cross-validation) evaluation results are presented.

4. Conclusion

Although very good ($\sim 90\%$ and more) performance is achieved on all 4 tasks using the kNN and the boosted decision tree classifiers, two important challenges defined by the benchmark remain, where more complex approaches in future work should improve the performance. On the one hand, by increasing the number of activities to be recognized — while keeping the same sensor set — the difficulty of the task exceeds the potential of standard methods (cf. the 'background' and 'all' tasks). On the other hand, when comparing classification performance individually for the 9 subjects, a high variance can be observed: the individual performance varies e.g. on the 'all' task between 74.02% and 100%. Therefore, personalization approaches (subject dependent training) could significantly improve on the results of the benchmark, and are highly encouraged.

Acknowledgements

This work has been performed within the project PAMAP (Physical Activity Monitoring for Aging People), funded under the AAL Joint Programme (AAL-2008-1). The authors would like to thank the project partners, the EU and national funding authorities for the financial support.

References

[1] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The WEKA Data Mining Software: an Update. *SIGKDD Explorations*, 11(1), 2009.

[2] A. Reiss and D. Stricker. Towards Global Aerobic Activity Monitoring. In *4th International Conference on Pervasive Technologies Related to Assistive Environments (PETRA)*, 2011.

[3] D. Roggen, S. Magnenat, M. Waibel, and G. Tröster. Wearable Computing: Designing and Sharing Activity Recognition Systems Across Platforms. *IEEE Robotics & Automation Magazine*, 18(2):83–95, 2011.