Feature Space Reduction for Multimodal Human Activity Recognition

Yale Hartmann, Hui Liu and Tanja Schultz
Cognitive Systems Lab, University of Bremen, Germany

Keywords: Human Activity Recognition, Biosensors, Multi-channel Signal Processing, Feature Space Reduction, Stacking.

Abstract: This work describes the implementation, optimization, and evaluation of a *Human Activity Recognition* (*HAR*) system using 21-channel biosignals. These biosignals capture multiple modalities, such as motion and muscle activity based on two 3D-inertial sensors, one 2D-goniometer, and four electromyographic sensors. We start with an early fusion, *HMM*-based recognition system which discriminates 18 human activities at 91% recognition accuracy. We then optimize preprocessing with a feature space reduction and feature vector stacking. For this purpose, a *Linear Discriminant Analysis* (*LDA*) was performed based on *HMM* state alignments. Our experimental results show that *LDA* feature space reduction improves recognition accuracy by four percentage points while stacking feature vectors currently does not show any positive effects. To the best of our knowledge, this is the first work on feature space reduction in a *HAR* system using various biosensors integrated into a knee bandage recognizing a diverse set of activities.

1 INTRODUCTION

Arthrosis is the most common joint disease worldwide and causes a noticeable reduction in quality of life. The research and development efforts to supporting arthrosis patients have increased significantly over the last years. Efforts range from active motion support systems via exoskeletons (Fleischer and Reinicke, 2005)(Liu et al., 2017), novel markers and methods for arthrosis diagnostics (Mezghani et al., 2017) to automatic offline and online recognition of human activities associated with causing strain on the knee (Rebelo et al., 2013)(Liu and Schultz, 2018)(Liu and Schultz, 2019). Additionally, *HAR* systems have been developed to detect and take action against functional decline based on the Stair Climb Power Test (Hellmers et al., 2017).

The mentioned works typically achieve recognition accuracies in the high ninety percent. (93% (Hellmers et al., 2017), 97% (Liu and Schultz, 2018) and 98% (Rebelo et al., 2013)). While these are great results, the scope of these studies is very specific, usually recognizing a couple of different activities and using similar sensor types. Sensors typically found in *HAR* Systems based around the knee and also *HAR* Systems, in general, include Accelerometers, Gyroscopes, (Bi-polar) Electromyography sensors and Electrogoniometers. Less typical setups have included Magnetometers and Barometers (Hellmers et al., 2017) or Piezoelectric and Airborne microphones (Teague et al., 2016)(Lukowicz et al., 2004).

Our goal is to technically assist the early treatment of arthrosis using a *HAR* system to measure and reflect on knee straining behavior. Furthermore, we aim to widen the scope by introducing a set of different and diverse activities and by contributing a base system to evaluate the benefit of sensors and features to the discrimination of these activities. In this paper, we will focus on that base system.

To achieve this goal we are following up on our work using biosensors integrated into a knee bandage (Liu and Schultz, 2018) and are using the same framework for data acquisition and annotation, but with a larger set of activities, sensors and newly evaluated parameters. We continue to model activities using *Hidden Markov Models* (*HMMs*), which is a widely adopted approach. Examples include the recognition of assembly and maintenance tasks (Lukowicz et al., 2004) or 3D handwriting recognition (Amma et al., 2010). However, differently, to most *HAR* systems we need to model very short activities using small windows, resulting in parameters similar to those found in *HMM*-based *Automatic Speech Recognition* (*ASR*) systems.

Hartmann, Y., Liu, H. and Schultz, T.
Feature Space Reduction for Multimodal Human Activity Recognition.
DOI: 10.5220/0008851401350140
In *Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2020) - Volume 4: BIOSIGNALS*, pages 135-140
ISBN: 978-989-758-398-8; ISSN: 2184-4305

2 EQUIPMENT AND SETUP

We chose the *biosignalsplux* Research Kits[1] as a recording device. One *PLUX* hub can process signals from 8 channels (each up to 16 bits) simultaneously. Therefore, we used three hubs connected via a cable to ensure synchronization during the entire session.

Similar to (Mathie et al., 2003) and (Liu and Schultz, 2018), we used two tri-axial accelerometers, four bipolar EMG sensors and both channels of one bi-axial electrogoniometer, as they were proven to be effective and efficient. Additionally, combined with the accelerometer into two *Inertial Measurement Units* (*IMU*), two tri-axial gyroscopes were used. We used both channels of an electrogoniometer to measure both the frontal and sagittal plane since we intend to recognize rotational movements of the knee joint in activities like ”curve-left” and ”curve-right”. Moreover, we used a piezo and airborne microphone like (Teague et al., 2016) and (Lukowicz et al., 2004) and included an additional force sensor.

EMG signals, as well as the microphones, require a high sampling rate, whereas the other biosignals are slow in nature. Therefore, they were recorded with different sampling rates. The slower signals used 100Hz and were up-sampled to match the 1000Hz used for the faster signals.

2.1 Sensor Placement

We use the Bauerfeind GenuTrain knee bandage[2] shown in Figure 1 as the wearable carrier of the biosensors. Table 1 lists all measured muscles and sensor positions. The sensor positioning was decided in collaboration with kinesiologists of the Institute of Sport and Sports Science at Karlsruhe Institute of Technology and designed to capture human everyday and sports activities relevant to gonarthrosis.

Table 1: Sensor placement and captured muscles.

Sensor	Position / Muscle
IMU1	Thigh, proximal ventral
IMU2	Shank, distal ventral
EMG1	Musculus vastus medialis
EMG2	Musculus tibialis anterior
EMG3	Musculus biceps femoris
EMG4	Musculus gastrocnemius
Goniometer	Right Knee, lateral
Microphones	Bandage inside/outside, medial
Force Sensor	Between patella and bandage

[1]biosignalsplux.com/researcher

[2]www.bauerfeind.de/en/products/supports-orthoses/knee-hip-thigh/genutrain.html

Figure 1: The knee bandage used as carrier.

3 DATASET

We recorded a dataset of eighteen activities from seven male subjects in a controlled lab environment using the previously introduced *ASK* framework (Liu and Schultz, 2018). We had to drop three of the seven subjects' recordings due to technical issues, resulting in a total of 40 minutes of usable semi-automatically annotated data. While this is a limited amount, further recordings are underway.

Table 2 gives occurrences, minimum and maximum length of the eighteen activities.

Most activities in Table 2 are self-explanatory, the others are defined thusly:

Curve-X-Spin is a fast 90° body turn in one step.

Curve-X-Step is a 90° turn using several walking steps.

V-Cut-X is a direction change with an acute angle at jogging speed.

Lateral-shuffle-X are repeated lateral steps starting with the right/left foot, the other following.

Jump-one-leg means jumping 1m forward using the bandaged leg.

Jump-two-legs means jumping in place using both legs.

Run means several steps passed at constant jogging speed.

An imbalance of occurrences can be observed for the activities ”Run” and ”Walk” and is explained by their repeated use in different lists, which we used similarly to (Liu and Schultz, 2018). However, this imbalance is welcomed as it reflects expectations in uncontrolled settings as well as allowing for a more detailed model better discriminating similar activities.

Table 2: Number of occurrences, minimum and maximum length of each activity.

Activity	Occ.	Min	Max.
curve-left-spin	54	0.725s	3.249s
curve-left-step	44	1.759s	4.269s
curve-right-spin	47	0.666s	2.279s
curve-right-step	45	1.649s	3.789s
jump-one-leg	47	0.749s	2.159s
jump-two-leg	47	0.999s	1.969s
lateral-shuffle-left	47	1.129s	4.089s
lateral-shuffle-right	47	0.969s	4.379s
run	86	1.179s	3.139s
sit	38	1.329s	5.089s
sit-to-stand	42	0.939s	3.589s
stair-down	45	1.769s	5.259s
stair-up	43	1.989s	5.159s
stand	37	1.759s	5.129s
stand-to-sit	42	1.029s	3.449s
v-cut-left	43	0.709s	2.209s
v-cut-right	37	0.679s	1.699s
walk	198	1.579s	5.179s

4 BASELINE HUMAN ACTIVITY RECOGNIZER

We developed a baseline recognizer using our in-house *HMM* Decoder *BioKIT* with simple features and tuned the parameters to achieve a good baseline. Our recognizer uses a forward topology commonly found in *HAR* and *ASR*, where each state allows for a transition to itself or the next(Rebelo et al., 2013). The emission probability for each state is modeled using *Gaussian Mixture Models* (*GMMs*). For our base system, we model each activity with the same number of states and each state with the same number of Gaussians per mixture.

4.1 Windowing and Feature Extraction

Feature Extraction is straightforward. For each channel, we use a rectangular window function with some overlap, then calculate the *Root Mean Square* (*RMS*) and Average (avg) on each window and z-normalize the whole activity. Due to the lack of spectral features, no smoothing window function is required.

Denoting the sample sequence of a window as $(x_1, \ldots x_n)$ and $N = n$ the number of samples in that window, the average is defined as:

$$avg = \frac{1}{N}\sum_{n=1}^{N} x_n, \tag{1}$$

The *Root Mean Square* is defined as:

$$RMS = \sqrt{\frac{1}{N}\sum_{k=1}^{N} x_n^2} \tag{2}$$

While our framework supports both early and late fusion of biosignals, we opted to use an early fusion in order to allow feature combinations of different channels. The resulting multi-channel feature vector has 42 dimensions since there are 21 channels and per channel two features are calculated. This approach differs from our previous work (Liu and Schultz, 2018) where *RMS* was only used for the EMG channels and Average for all others.

4.2 Parameter Tuning

The research on gait analysis commonly distinguishes two phases into eight events, as described in (Whittle, 2014) and further discussed in (Whittle, 1996) and (Mezghani et al., 2013). Therefore, we expected a *HMM* topology with eight states to perform best. However, in evaluation it was outperformed by a six-state topology, which is what we then continued with. This unexpected result might be due to the *GMMS* being better fitted in the six-state topology or due to not all activities corresponding to a gait cycle. Further discussion is left to future work.

The overall shortest activity is an instance of ”curve-right-spin” (See Table 2) at 666ms. Therefore, requiring a maximum window length of 111ms to be recognizable with a six-state *HMM*. Assuming an equal distribution of samples across the *HMM* when aligned, we opted to use a window length of 10ms with an overlap of 2ms, resulting in an absolute minimum of twelve samples per state. Using these parameters we evaluated several numbers of Gaussians per mixture, finding that seven Gaussians perform reasonably well and having enough data to fit the Gaussians properly. In the future, we want to use a merge and split estimation here, adaptively adjusting the amount of Gaussians to use the present data optimally.

To summarize, our base system used a multi-channel feature signal rectangularly framed with a window length of 10ms and an overlap of 2ms. Calculating the Average and *RMS* on each window and z-normalizing the result. *HMMs* are modeled using six states with a forward topology and each state using a *GMM* with seven Gaussians. Evaluating this setup with a randomized ten-fold cross-validation the mean accuracy is 91%.

5 FEATURE SPACE REDUCTION

Reducing the feature space dimension has several benefits. Firstly, *GMMs* can be fitted more effectively. Secondly, more (sensor specific) features can be added easily, since the *LDA* is used to transform them into a smaller maximum discriminating feature space. Effectively reducing redundancy between channels and producing a consistent feature space dimension independent from sensor and feature setup.

Reducing the original multi-channel feature vector directly with a *LDA* could not preserve non-linear relations and would interfere with the sequential modeling of our *HMMs*. Therefore, we align the samples to states using the Viterbi algorithm and using each activities' state as target for the aligned feature vector. This approach allows for a linear supervised reducing function, which generally outperform unsupervised options like a *Principal Component Analysis* (*PCA*). Similar experiments using non-linear reduction methods like *Neural Networks* (*NN*) combined with a *PCA* have shown to improve performance(Hu and Zahorian, 2010).

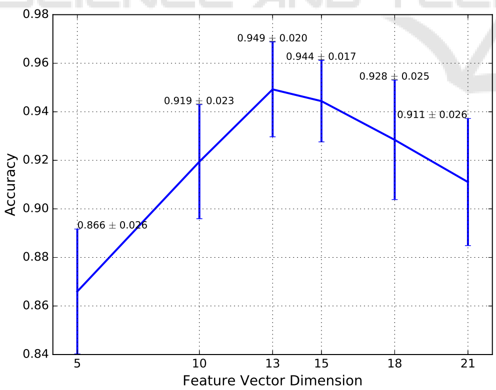

Figure 2: Evaluation: Feature Dimension. Setup: 10ms window length; 2ms overlap; 6 States per *HMM*; 7 Gaussians per state.

Applying this technique to our base system, the system has a performance peak when reducing to a 13-dimensional feature space as shown in Figure 2. Increasing the accuracy by four percentage points to 94.9% compared to the approach without dimension reduction. Similarly clear is the decrease in performance when reducing to too few dimensions as too few meaningful features are used. More surprising is the steady decline of performance after 13 dimensions. There are several possible reasons for this, including too high a dimension to fit the *Gaussians* properly as well as sensors that might provide contradictory information. A deeper analysis of sensors and features is planned for the future.

The recognizer using 13-dimensional features tends to predict walking and running over curve based activities as seen in Figure 3. Several Steps and Spins are recognized as "Walk". The "V-Cuts" are more than once predicted to be "Run". Additionally, the recognizer confuses left and right "V-Cuts". This is consistent behavior to what we have seen in the baseline recognizer.

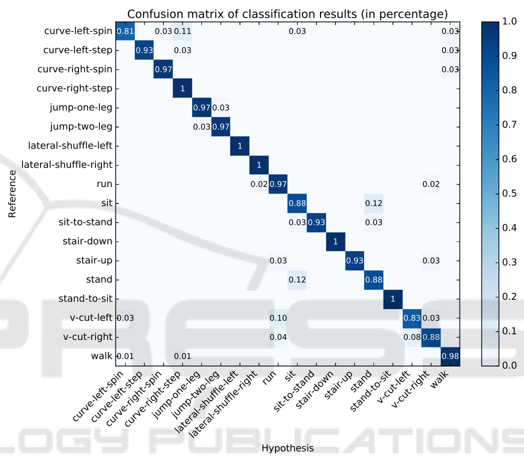

Figure 3: Confusion matrix of recognition results. Setup: 10ms window length; 2ms overlap; 6 States per *HMM*; 7 Gaussians per state; 13-dimensional feature space.

6 FEATURE VECTOR STACKING

Another possible approach to improve performance is to add context to each feature vector in the preprocessing step by prepending the *n* previous and appending the *n* following feature vectors. This process is called *stacking*.

If evaluated naively without a feature space reduction, the performance decreases with increasing context. This behavior is expected due to the significantly increasing feature vector dimensions as shown in Table 3 compared to only few data samples.

Running the same experiment with a fixed feature space dimension of 13 does not increase the performance (Figure 4). On the contrary, 0-stacking features enhance performance in our case. Incidentally, 0-stacking here uses the same configuration as the recognizer shown in Figure 3. Additionally, the per-

Table 3: Results of naive *stacking*. Window length: 10ms; overlap: 2ms; 6 States per *HMM*; 7 Gaussians per state; without feature space reduction.

Context	Accuracy	Vector Dimension
0	0.92	42
1	0.80	126
2	0.74	210

formance between different context sizes is not significant as a statistical analysis via T-Test indicates.

These results are obtained on a local optimum with a 13-dimensional feature space using solely statistical features. Therefore, we will investigate this behavior for temporal features and further dimensions in the future.

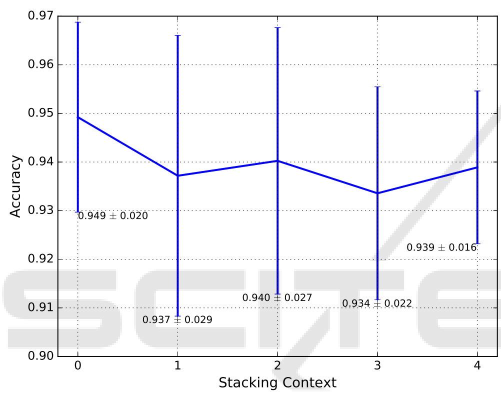

Figure 4: Evaluation: *stacking*. Setup: 10ms window length; 2ms overlap; 6 States per *HMM*; 7 Gaussians per state; 13-dimensional feature space.

7 CONCLUSION

In this paper, we successfully implemented, evaluated and improved an offline, early fusion *HAR* system using a 21-dimensional biosignal comprised of different sensors placed onto a knee bandage. The base system performed very well with a 91% accuracy using only simple features. We showed, that the performance could be improved by four percentage points to 94.9% using a *LDA* trained with *HMM* state aligned labeled data and reducing the feature space dimension. Furthermore, we found that in our case stacking feature vectors to improve context did not increase performance but instead slightly decreased it with respect to not stacking at all, which is one topic for further investigation.

In the future, we will evaluate additional more sophisticated features targeted to the specific sensors and their influence on the overall performance as well as the feature space reduction specific performance. Furthermore, we will create and evaluate different topologies for different activities and investigate the performance of our system using a person independent evaluation on a larger dataset. To the best of our knowledge, this is the first work on feature space reduction in a *HAR* system using various biosensors integrated into a knee bandage recognizing a diverse set of activities.

REFERENCES

Amma, C., Gehrig, D., and Schultz, T. (2010). Airwriting recognition using wearable motion sensors. In *First Augmented Human International Conference*, page 10. ACM.

Fleischer, C. and Reinicke, C. (2005). Predicting the intended motion with emg signals for an exoskeleton orthosis controller. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2005)*, pages 2029–2034.

Hellmers, S., Kromke, T., Dasenbrock, L., Heinks, A., Bauer, J. r. M., Hein, A., and Fudickar, S. (2017). Stair Climb Power Measurements via Inertial Measurement Units. pages 1–9.

Hu, H. and Zahorian, S. A. (2010). Dimensionality reduction methods for HMM phonetic recognition. In *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 4854–4857. IEEE.

Liu, H. and Schultz, T. (2018). Ask: A framework for data acquisition and activity recognition. In *11th International Conference on Bio-inspired Systems and Signal Processing, Madeira, Portugal*, pages 262–268.

Liu, H. and Schultz, T. (2019). A Wearable Real-time Human Activity Recognition System using Biosensors Integrated into a Knee Bandage. In *Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies*, pages 47–55. SCITEPRESS - Science and Technology Publications.

Liu, X., Zhou, Z., Mai, J., and Wang, Q. (2017). Multi-class SVM Based Real-Time Recognition of Sit-to-Stand and Stand-to-Sit Transitions for a Bionic Knee Exoskeleton in Transparent Mode. In *The Semantic Web - ISWC 2015*, pages 262–272. Springer International Publishing, Cham.

Lukowicz, P., Ward, J. A., Junker, H., Stäger, M., Tröster, G., Atrash, A., and Starner, T. (2004). Recognizing Workshop Activity Using Body Worn Microphones and Accelerometers. In *Pervasive Computing*, pages 18–32. Springer, Berlin, Heidelberg, Berlin, Heidelberg.

Mathie, M., Coster, A., Lovell, N., and Celler, B. (2003). Detection of daily physical activities using a triaxial accelerometer. In *Medical and Biological Engineering and Computing*. 41(3):296—301.

Mezghani, N., Fuentes, A., Gaudreault, N., Mitiche, A., Aissaoui, R., Hagmeister, N., and De Guise, J. A.

(2013). Identification of knee frontal plane kinematic patterns in normal gait by principal component analysis. *Journal of Mechanics in Medicine and Biology*, 13(3).

Mezghani, N., Ouakrim, Y., Fuentes, A., Mitiche, A., Hagemeister, N., Vendittoli, P.-A., and De Guise, J. A. (2017). Mechanical biomarkers of medial compartment knee osteoarthritis diagnosis and severity grading: Discovery phase. *Journal of Biomechanics*, 52:106–112.

Rebelo, D., Amma, C., Gamboa, H., and Schultz, T. (2013). Activity recognition for an intelligent knee orthosis. In *6th International Conference on Bio-inspired Systems and Signal Processing*, pages 368–371. BIOSIGNALS 2013.

Teague, C. N., Hersek, S., Toreyin, H., Millard-Stafford, M. L., Jones, M. L., Kogler, G. F., Sawka, M. N., and Inan, O. T. (2016). Novel Methods for Sensing Acoustical Emissions From the Knee for Wearable Joint Health Assessment. *IEEE Transactions on Biomedical Engineering*, 63(8):1581–1590.

Whittle, M. W. (1996). Clinical gait analysis: A review. *Human Movement Science*, 15(3):369–387.

Whittle, M. W. (2014). *Gait Analysis*. Normal Gait. Butterworth-Heinemann.