21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

# Unsupervised Motion Pattern Learning for Motion Segmentation

Markus Weber, Gabriele Bleser, Marcus Liwicki, Didier Stricker
*German Research Center for AI (DFKI GmbH)*
*Kaiserslautern, Germany*
*{firstname.lastname}@dfki.de*

## Abstract

*This paper proposes a novel method for automated generation of motion segmentation models for full body motion monitoring. The method generates, in an unsupervised manner, a motion template for a dynamic warping approach from a short training sequence,* i.e., *from very few data. Therefore it first automatically detects motif candidates,* i.e. *the recurring patterns in the training sequence. Then it uses the detected motifs to construct the model. This novel method is able to automatically find motifs in a multivariate time series and generate a model which is capable of segmenting the series in a real-time system. The technology is evaluated in the context of a personalized virtual rehabilitation trainer application during a clinical study. The novel motion capturing dataset is publicly available.*

## 1 Introduction

Bodytracking becomes more and more popular as with the Microsoft Kinect an affordable sensor is now available on the consumer market. In medical applications bodytracking can be used for tracking the motion of patients and monitor their correct execution. The PAMAP (Physical Activity Monitoring for Aging People) project describes such a scenario where motion tracking data is recorded during a supervised session with a physiotherapist and used for automatic supervision of the patient at home. In the supervised training session each patient performs exercises together with a physician, who shows how to perform the exercises according to his or her special needs. In order to realize automatic monitoring at home, motion models are constructed during the training session and then used by the system to monitor the correct execution of exercises, automatically count the number of repetitions, evaluates segmented motion segments, and, based on this, gives immediate feedback to the patient [3].

In this paper, a fully automated method for detecting motion motifs is developed. Noteworthy, motion capture data is time series data, and time series have been studied in many scientific areas including chemistry, physics, geology, meteorology and economics. In biological science, systems such as MEME [1] were developed for discovering motifs in DNA and protein sequences. Usually, they are designed to work with categorical sequences, and thus only a few systems are applicable to real-valued time series analysis. Jensen et al. [4] generalized motif discovery over both categorical and continuous data and across arbitrary similarity metrics. Within the data mining community, Chiu et al. [2] developed an efficient, probabilistic algorithm for motif discovery using a form of locality-sensitive hashing. Another approach is based on discretization, which is performed by symbolic aggregate approximation (SAX) [6]. Although the original presentation only worked with univariate data, SAX can also be applied to multivariate time series by concatenating the words representing each dimension. The PERUSE algorithm discovers motifs directly in multivariate time series and allows both non-linear time warping and variable-length motifs [9]. It uses a model similar to a left-right HMM with an explicit, Gaussian distribution over the time between each sample.

Using a motif discovery approach a system is capable of extracting recurring patterns of a sequence and use this to train a system for segmentation of this specific pattern. Time series segmentation is important in many of these areas because it is often used as a preprocessing step in time series analysis applications. As the segmentation process is quite important for the performance of the overall system it has to be quite reliable. This paper is focusing on automatically finding motifs in a multivariant motion capturing data and generate a model which is capable of segmenting motion in a real-time system. An existing approach for solving this problem are based on a predefined window size [7].

We introduce a novel step to estimate a suitable window size based on the dominant frequency.

The rest of the paper is organized as follows Section 2 describes the proposed unsupervised method to automatically find motion patterns from very few data. In Section 3 the results of the clinical trials are presented. Finally, the work will be concluded in Section 4 with future work and a conclusion.

## 2 Motif Segmentation

Motion segmentation is one of the key techniques in the context of motion analysis. The basic idea is to split motion capture data into segments — begin of motion pattern and its end — which can be used to evaluate this motion sequence, for instance in a fitness trainer context, where the system gives feedback if the exercise is performed correctly.

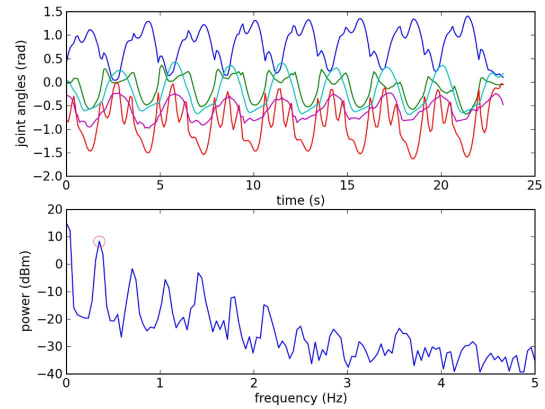
Note that motion capturing data is multidimensional and the number of dimensions is dependent on how the motion capture data is represented. In our scenario the full body is capture, using ten inertial measurement units (IMU) [10]. So, the complete model consists of ten rigid bodies (bones: torso, pelvis, upper arms, forearms, upper legs, lower legs) connected by anatomically motivated restricted joints. The rotations of torso and pelvis are modeled with three degrees of freedom (DOF), shoulder and hip joints a as ball joint, thus also three DOF. Only the arm and knee are modeled as pivot joints thus have two DOF, hence in a sum we have a signal with 26 dimensions.

Based on the joint angles provided by the motion capture system, this section describes a fully automated method for unsupervised discovery of motion patterns from a very short training sequence, such as the one shown in Figure 1a. The training data is assumed to contain a predefined number of pattern examples performed by the user during the training step.

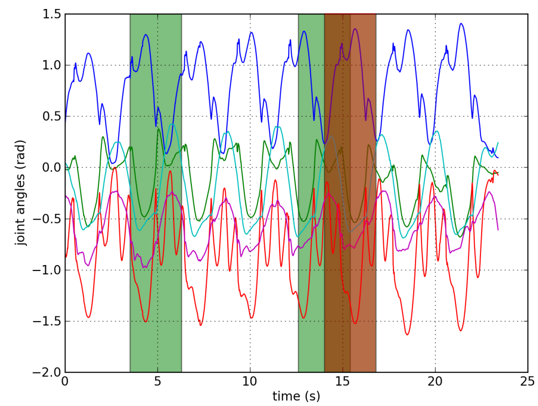
The proposed method for model generation first automatically detects motif candidates, *i.e.* the recurring patterns in the training sequence. Then it uses the detected motifs to construct the model.

In the following we will describe the several steps of motif discovery, *i.e.*, how the recurring patterns are detected. The first step of motif discovery is to reduce the dimensionality of the motion data. Based on the assumption that the most moving joints contain the most relevant information, we simply detect the channels with highest variance. Therefore, we compute mean value of the variances for all channels of the time series — a channel is an Euler component of a joint. Let

$C_i = \{(t_1, c_1), (t_2, c_{2,1}), \cdots (t_n, c_n)\}$, with $0 \leq i < b$




(a) Combined PSD of all channels of the reduced body tracking signal. The dominant frequency is marked with a red circle.




(b) Motif candidates. The selected motif is marked in red.

Figure 1: Example for the different steps of the motif discovery on a training sequence.

be a time series for a joint dimension $i$ of the motion, where $b$ is the number of tracked joint dimensions and $n$ the lenght of the sequence. Hence, each channel where $Var(C_i) > \mu_{var}$ is considered to be relevant, with

$$\mu_{var} = \frac{1}{b} \cdot \sum_{i=0}^{i<b} Var(C_i).$$

The second step is to estimate a suitable window size. As mentioned before, existing approaches are based on a predefined window size [7, 8]. Since in our case the length of the pattern is unknown, we have to estimate a suitable window size $w_{est}$. For the most moving joints we have observed a periodic signal behavior, as the exercises are performed Based on the assumption that the repetitions in the training sequence are performed consecutively with roughly the same speed, a dominant frequency should be present in the signal. This can be extracted using the combined power spectral density (PSD) [12] (*cf.* Figure 1a). The window length $w_{est}$ is then initialized as the wavelength of the dominant frequency, $w_{est} = \lambda = \frac{v}{f_{dominant}}$, with $v$ being the sampling rate.

The next step detects the motif candidates. Therefore an extended version of Minnen's method [7] parametrized with $w_{est}$ is used. The method collects overlapping subsequences, $S_i$, of length $w_{est}$ from the training signal, $S$, and determines the $k$-nearest neighbors for each subsequence as $\mathrm{kNN}(S_i) = S_{i,1..k}$. Here, $k$ is the predefined number of repetitions. In order to reduce the sensitivity to local time shift and slightly varying execution speed, dynamic time warping (DTW) is used as distance measure. A real motif should have at least $k$ similar subsequences. Hence, in order to find good motif candidates, for each subsequence, $S_i$, the distance density is estimated as the reciprocal of the distance to the least similar neighbor $k$: $den(S_i) \propto \frac{1}{dist(S_i,S_{i,k})}$ The motif candidates, $cand_i$, are then identified as the local maxima of the densities among its $k$ nearest neighbors:

$$maxima(S_i) = S_i : \forall\, S_{i,j}\ den(S_i) > den(S_{i,j}),$$

where $j = [1, k]$. Motif candidates are highlighted in Fig. 1b.

In the next step of the algorithm a model for each candidate is generated and used to segment the signal. As most of the learning approaches fail if there is only few training samples are available, either constructed models [10] or template-based approaches are considerable. We have chosen a template-based approach, based on the Online DTW proposed in [5]. The motif candidate is chosen as the template for the DTW and its k-Neighbors to choose threshold for the costs.

In the final step of the algorithm the candidate, which model segments the signal best, is chosen as the motion motif. As the criterion, the difference between the segmented patterns and the known number of executions and the mean of the normalized DTW costs of the segments. The selected candidate for our example is marked red in Fig. 1b.

The chosen motif and its nearest neighbors can now be either used to generate the class template from a set of the best templates, *e.g.*, extract the templates from the nearest neighbors that have the best minimum interclass DTW distances [5] or generate a Hidden Markov Model as proposed in [10]. Both approaches are suitable for an online real-time segmentation.

## 3 Evaluation

The proposed method was evaluated in the clinical study of the PAMAP project. 30 elderly patient, in the age of 55 to 86 years, participated in the trials. Three different groups of patients took part in the study (each 10 patients), healthy patients, cardiac patients and functional disabled subjects. For each group a specific training plan with 10 to 13 different exercises has been performed by the patients [1].

For capturing the patients motion, a reliable and accurate sensor fusion approach based on miniature body-mounted inertial measurement units (IMUs) has been developed [11]. Under full operation, the whole body can be captured with ten IMUs. The captured motion signals during training are then input for the proposed motif discovery method.

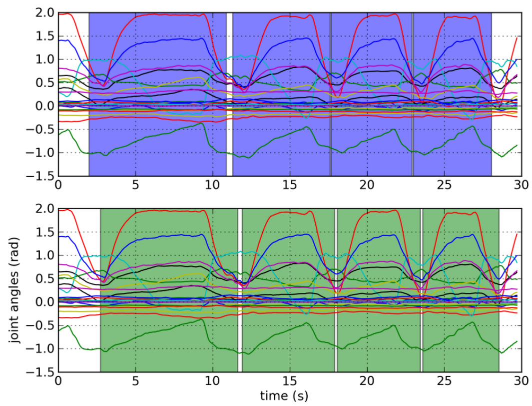
In order to evaluate the performance of the method we determined the precision and recall, as well as an overlap for the segmented signal. Precision is the fraction of segmented sequences that are relevant which means execution of the performed exercises, while recall is the fraction of correct retrieved executions. A segment is considered a corretly retrieved if it at least overlaps with the ground truth segment with 30%. The overlap is the percentage of overlapping of the segmented motion segment with the ground truth segment.

Table 1 lists the results of experiments for each patient (precision, recall, and overlap). Figure 2a shows an example of promising result for a biceps exercise, whereas Figure 2b illustrates the most common error, that the middle of the exercise has been assumed to be the beginning. Hence, the exercise will be correctly recognized in precision and recall, as only the last execution is not detected, because the end position of the shifted model is the middle point of the actual exercise which results in an overlap of $\sim 0.5$. For the future work, knowledge about feasible beginning poses can be used to overcome this problem.

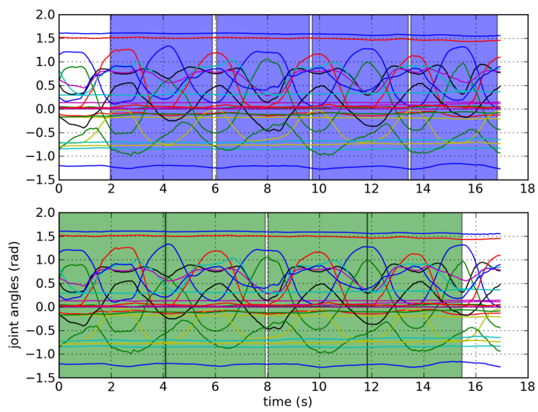
Table 1: Results per patient. Precision, recall and percentual overlap.

| # | **Precision** $\mu \pm var$ | **Recall** $\mu \pm var$ | **Overlap** $\mu \pm var$ |
|---|---|---|---|
| 1 | 0.77±0.10 | 0.80±0.10 | 0.59±0.06 |
| 2 | 0.85±0.08 | 0.86±0.08 | 0.71±0.08 |
| 3 | 0.98±0.00 | 0.97±0.01 | 0.75±0.03 |
| 4 | 0.93±0.02 | 0.95±0.02 | 0.76±0.02 |
| 5 | 0.98±0.00 | 0.90±0.01 | 0.73±0.03 |
| ... | | ... | |
| 30 | 0.95±0.01 | 0.88±0.01 | 0.59±0.04 |
| ⌀ | 0.93±0.00 | 0.91±0.00 | 0.72±0.00 |

[1] The motion capturing dataset is publicly available at: `http://www.pamap.org/PAMAP_trials.tar.gz`




(a) Biceps curl. Upper plot shows results of the motif discovery (blue), and the lower plot the ground truth (green).




(b) Thoraric opening, bend back. Shifted beginning error.

Figure 2: Results of the segmentation using the model of the selected motif candidate.

## 4 Conclusion and future work

This paper presents a fully automated method for detecting motion motifs is developed. In order to deal with this real-time aspects of online motion monitoring, we have chosen the Online DTW [5] which is capable of dealing with a 100 Hz signal. The system has been evaluated on a new publicly available database.

Note that the proposed methods are work in progress, but the results of the clinical trial are promising. Our unsupervised approach can be used to automatically learn a segmentation model from a short training sequence. Future work will focus on methods to incorporate knowledge about human movement to reject implausible beginning poses.

## References


[1] T. Bailey and C. Elkan. Fitting a mixture model by expectation maximization to discover motifs in biopolymers, 1994.

[2] B. Chiu, E. Keogh, and S. Lonardi. Probabilistic discovery of time series motifs. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '03, pages 493–498, New York, NY, USA, 2003. ACM.

[3] L. Fradet and F. Marin. Strength exercise monitoring: what to evaluate and how? *Computer Methods in Biomechanics and Biomedical Engineering*, 14(sup1):35–36, 2011.

[4] K. L. Jensen, M. P. Styczynski, I. Rigoutsos, and G. N. Stephanopoulos. A generic motif discovery algorithm for sequential data. *Bioinformatics*, 22(1):21–28, 2006.

[5] M. H. Ko, G. West, S. Venkatesh, and M. Kumar. Using dynamic time warping for online temporal fusion in multisensor systems. *Information Fusion*, pages 370 – 388, 2008.

[6] J. Lin, E. Keogh, S. Lonardi, and B. Chiu. A symbolic representation of time series, with implications for streaming algorithms. In *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*, DMKD '03, pages 2–11, New York, NY, USA, 2003. ACM.

[7] D. Minnen, C. L. Isbell, I. Essa, and T. Starner. Discovering multivariate motifs using subsequence density estimation and greedy mixture learning. In *Proceedings of the 22nd national conference on Artificial intelligence - Volume 1*, pages 615–620. AAAI Press, 2007.

[8] D. Minnen, T. Starner, I. Essa, and C. Isbell. Discovering Characteristic Actions from On-Body Sensor Data. *2006 10th IEEE International Symposium on Wearable Computers*, pages 11–18, Oct. 2006.

[9] T. Oates. Peruse: An unsupervised algorithm for finding recurrig patterns in time series. In *Proceedings of the 2002 IEEE International Conference on Data Mining*, ICDM '02, pages 330–, Washington, DC, USA, 2002. IEEE Computer Society.

[10] M. Weber, G. Bleser, G. Hendeby, A. Reiss, and D. Stricker. Unsupervised Model Generation for Motion Monitoring. In *IEEE International Conference on Systems, Man and Cybernetics - Workshop on Robust Machine Learning Techniques for Human Activity Recognition*, pages 51–54, Anchorage, 2011. IEEE.

[11] M. Weber, G. Bleser, G. Hendeby, A. Reiss, and D. Stricker. Unsupervised model generation for motion monitoring. In *IEEE SMC 2011.Workshop on Robust Machine Learning Techniques for Human Activity Recognition.*, pages 51–54. IEEE, Anchorage, 10 2011.

[12] P. Welch. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15(2):70–73, 1967.