

Wavelet Statistics for Human Motion Classification

Kevin Quennesson

GVU Center
Georgia Institute of Technology
Atlanta, GA

Elias Ioup

Naval Research Laboratory
Stennis Space Center, MS

Charles Isbell

GVU Center
Georgia Institute of Technology
Atlanta, GA

Abstract

Human motion is as much characterized by its low frequency shape as by its high frequency temporal discontinuities – such as when a joint reaches its physical limit or when a foot touches the floor. Wavelets are particularly efficient at capturing both high and low frequency information. We introduce a method of classifying human motion using wavelet coefficients to build a representation of human motion signals. The representation is computed by finding the histograms of the wavelet coefficients previously scaled according to frequency. We use Support Vector Machines (SVMs) to classify those histograms and demonstrate the accuracy of the method on human motion gathered from both a motion capture systems and accelerometers.

Introduction

Human motion data is becoming increasingly accessible as motion capture systems become more affordable and other sensing devices more robust. A challenge is to classify this large amount of data. We are interested in classifying different styles of human motion data efficiently and automatically. Existing methods mostly consider the spatial configuration of signals, which makes it hard to differentiate motions which look spatially the same but have different frequential characteristics – for instance walks of different speeds.

Our method differentiate signals, not on spatial, but spatio-temporal differences. It is based on characterizing signals by computing statistics on their wavelet coefficients at all frequency bands. We regroup those statistics in one vector on which a SVM classifier is trained.

Previous approaches classify motions based strictly on their spatial configuration. (Barbic *et al.* 2004) uses PCA, while Support Vector Machines (SVM) are used in (Ramanan & Forsyth 2003) on frames of video data and in (Arikan & Forsyth 2002) on frames of motion capture data. Ignoring the frequential component of a motion leads to an undesirable class of systemic error. For example, (Ramanan & Forsyth 2003) cites misclassifications between walking and running motions. In response, (Tanawongsuwan & Bobick 2001) have integrated a scale component in the Dynamic Time Warping (DTW); however, time is consider

aposteriori along a window of motion. Similarly, this occurs if measuring same-state transitions for HMM. We would like to avoid such systemic error by using information on space and frequency simultaneously.

Histograms of wavelet coefficients

Properties of the wavelet transform

We present here some properties of the wavelet transform which makes it well suited for capturing both low frequency shapes and high frequency temporal discontinuities of signals.

The wavelet transform of a signal f at point u in time and s in frequency is defined by the convolution of f with the wavelet $\psi_{u,s}$: $Wf(u, s) = f \star \psi_{u,s}$. The wavelet $\psi_{u,s}(t) = \frac{1}{\sqrt{s}} \psi(\frac{t-u}{s})$ is the translation by u followed by the dilation by s of an original wavelet ψ . One might see from computing the Fourier transform of $Wf(u, s)$ than it analyzes the signal in a “box” (called the Heisenberg box) centered around u in time and s in frequency. This “box” width (size along the temporal axis) decreases as the frequency s increases, whereas its height (size along the frequential axis) increases (since $width \times height$ is constant). A coefficient corresponding to a higher frequency will consequently be studied in a shorter period in time, which is more adapted to its variations. A window Fourier transform has fixed “box” width and height. Since temporal discontinuities are characterized by many high frequency oscillations, a Fourier transform will need many coefficients to represent a signal with discontinuities, whereas the wavelet transform will not.

Histogram representation

A wavelet transform provides a set of coefficients at different frequency bands. It is hard to align those coefficients between different signals for classification. We instead incorporate all coefficients into a single compact representation which does not need alignment. This representation can then be used as input for training a SVM classifier.

Our representation consists of a histogram of coefficients at *all frequency bands*. However, in the case of human motion, low frequency variations have more influence on the signal than high frequency temporal discontinuities, which

Type	Contents	Number of instances	Misclassified instances	Classification accuracy (10-fold cross validation)
Motion capture	Walk, jog, jump, slow walk, quick walk, monkey walk, march	38 Walk (9), jog(10), jump(5), slow walk(4), quick walk(4), monkey walk(2), march(4)	1	97.3%
Motion capture	Same as above without march	34	0	100%
Motion capture	Split long motion of walk, jump, jog	15	1	93.3%
Accelerometer and gyroscop data	6 types of human exercise: flat curl, shoulder ext., back, twist curl, shoulder press	191 (about 30 for each exercise)	7	96.3%

Table 1: Results on 4 data sets

implies that magnitudes of wavelet coefficients are smaller in high frequency bands than in low frequency bands. It will be hard for a histogram to capture all that information. For instance, most of the high frequency coefficients will be in the zero bin of a histogram adapted to low frequency band coefficients. To counter that problem we scale down the coefficients by a constant C as the frequency band decreases, so that most of them are contained in the $[-1, 1]$ range. We found that $C = 4$ gives satisfying results. Then the histograms are computed over the scaled coefficients. A histogram bin centered at x describes the importance of variations at all frequency bands of a scaled magnitude x . We show a human tibia joint angle signal, its wavelet decomposition and histogram representation in Figure 1.

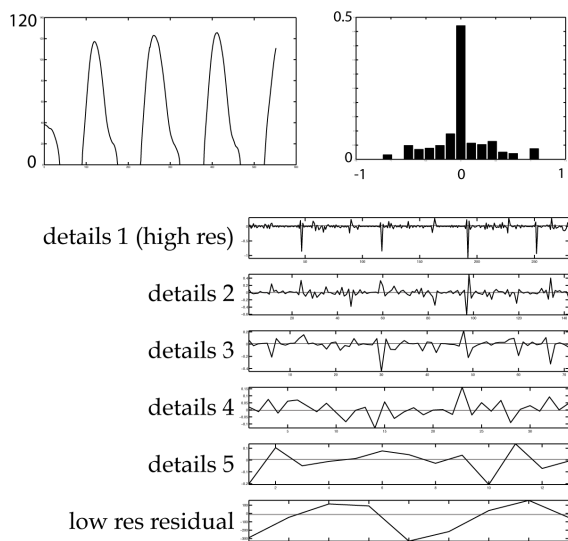


Figure 1: Left to right and top to bottom: signal of tibia joint angle (direction parallel to motion) for a marching motion, its histogram representation and its wavelet decomposition. Marching produces symmetric signals, which can be seen on the histogram, approximately symmetric around $x = 0$.

Human motions are often described with multiple dimensions. We build a vector by concatenating the histograms corresponding to the different dimensions of the signals. We

then train a SVM classifier on that vector. One motion is then characterized by one vector. The size of the vector is the same for all motions, even if their lengths are different. We provide results in the next section.

Results and discussion

Our results are presented in Table 1, all using 10-fold cross-validation. We found out experimentally that a bin size of 0.15 gives us the best results. We used biorthogonal wavelets with 6 vanishing points, which give a very sparse representation and only shows significant variations at the different frequency bands. Our dyadic wavelet transform is of level 8 so that no information remains in the continuous component of the decomposition.

We observe that our classifier efficiently distinguishes motions with different rhythms, such as quick walks, slow walks and normal walks.

Acknowledgments

The motion capture data used in this project was obtained from mocap.cs.cmu.edu. This database was created with funding from NSF EIA-0196217.

References

- Arikan, O., and Forsyth, D. A. 2002. Interactive motion generation from examples. In *SIGGRAPH '02: Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, 483–490. New York, NY, USA: ACM Press.
- Barbic, J.; Safonova, A.; Pan, J.-Y.; Faloutsos, C.; Hodgins, J. K.; and Pollard, N. S. 2004. Segmenting motion capture data into distinct behaviors. In *GI '04: Proceedings of the 2004 conference on Graphics interface*, 185–194. School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada: Canadian Human-Computer Communications Society.
- Ramanan, D., and Forsyth, D. A. 2003. Automatic annotation of everyday movements. In *NIPS*.
- Tanawongsuwan, R., and Bobick, A. F. 2001. Gait recognition from time-normalized joint-angle trajectories in the walking plane. In *CVPR (2)*, 726–731.