1-1-2008

Motion segmentation for humanoid control planning

Matthew Field
University of Wollongong, field@uow.edu.au

David A. Stirling
University of Wollongong, stirling@uow.edu.au

Fazel Naghdy
University of Wollongong, fazel@uow.edu.au

Zengxi Pan
University of Wollongong, zengxi@uow.edu.au

Motion segmentation for humanoid control planning

Matthew Field, David Stirling, Fazel Naghdy, Zengxi Pan
University of Wollongong, Australia
{mf91, stirling, fazel, zengxi}@uow.edu.au

Abstract

The discovery of major management behaviours from human motion data and uncovering their underlying components is investigated. A range of methods for segmenting major shifts in multidimensional time series are compared in inducing plausible behaviours from motion data. These behaviours are considered as supersets of motion primitives that define a repertoire of manoeuvres available to the human. The resulting multilayered symbolic model is used as a framework for humanoid imitation and control. It is hoped that with appropriate matching and scaling of degrees of freedom, models can be tested by extracting a trajectory for a simulation of the Nao soccer bot.

1 Introduction

Intelligent systems require flexible control skills in order to operate in the real world, but it is difficult to program this capability. A current trend in autonomous systems is in imitating behaviour from more complex agents, namely humans. The learning agent may then subsequently formulate successful action plans to control a system. In this sense it is a data driven approach to control, which facilitates adaptive learning mechanisms as opposed to optimal hard-coded actions.

The problem considered in this paper is learning control from human motion data and transferring the skills to a humanoid robot. Among the challenges presented by this domain are the ability to autonomously organise or categorise what is observed and extracting the intention or goal of the actions. Typically the data-driven models are not interpretable without observing example responses, we aim to find descriptive models that automatically differentiate between behaviours at a task level and 'proprioceptive' subgoals at a motor control or trajectory level.

1.1 Related Work

The data-driven techniques used in these learning tasks were first implemented in computer animation research where similar behaviours were clustered in a motion capture database. The most common techniques used in recent research involve encoding trajectories into Gaussian Mixture Models (GMM) or Hidden Markov Models (HMM). Other methods encode the high dimensional data into a low dimensional manifold using linear methods such as Principal Component Analysis (PCA) or Multi-dimensional Scaling (MDS). However, nonlinear mappings such as Isomap, Local Linear Embedding (LLE) or Gaussian Process Latent Variable Model (GPLVM) have shown higher performance in capturing relevant data structure.

The use of HMM in learning and prediction for multidimensional time series is widespread and data-driven robot imitation is no exception. [Inamura *et al.*, 2004] based their symbolic HMM approach on the mirror neuron hypothesis in primates. A number of HMM states were trained on motion capture sequences such that each state embodied a posture for the robot. States were compared in a 'proto-symbol' space and merged based on their relative Kullback-Leibler distances. The reproduction of motion trajectories from these states reflected the observation-motion duality of mirror neuron analogue. [Kulic *et al.*, 2008] expanded upon this framework by incrementally updating the model and creating a hierarchy of HMM sequences using Factorial HMM.

[Calinon *et al.*, 2007] clustered similar postures into a GMM of a size determined using the Bayesian Information Criterion (BIC). Generalized trajectories could be restored and reproduced in a humanoid in different contexts by using Gaussian Mixture Regression (GMR) between the appropriate sequence of states. The approach also relied upon a PCA pre-processing step which allows for fast online computation with the loss of some information.

Behaviour segmentation is also a recurring theme in computer animation research. [Beaudoin *et al.*, 2008]

mined motifs in large motion capture databases to create structured graphs which can blend fluid animations. Segmentation techniques were also evaluated by [Barbic *et al.*, 2004] for automating motion capture editing. Non-linear dimension reduction techniques as used by [Wang *et al.*, 2008] embed the data onto meaningful planes of motion style and content with relatively small data sets. Their methods based on GPLVM could interpolate in regions of the latent space where there were no observed data. This algorithm has also been implemented in humanoid imitation [Shon *et al.*, 2006] by projecting data from the latent space on the robots reduced DOF with some success.

If the models constructed are to be interpretable, symbolic learning techniques may form appropriate descriptions. This approach was used by [Tanaka *et al.*, 2005] in identifying sporting motifs and understanding their meaning. The data was reduced to one symbolic string and compressed using a Minimum Description Length (MDL) algorithm to find the most significant motifs. [Sun *et al.*, 2006] formulated a GMM of motion primitives from inertial data gathered from different hand movements. Further motions of the same style could be recognised using symbolic sequencing techniques.

2 Experimental Rig

The sensor system employed for human measurement is inertial motion capture equipment from Xsens Technologies, Moven. It has 16 inertial measurement units (IMU) in a stretch suit which locates sensors on major body landmarks to record corresponding angles. Assuming rigid body dynamics, the angles provide forward kinematics for a skeletal reconstruction of the human body. A single sensor has an orientation accuracy to 2° RMS, however, with the feedback from a chain of sensors there is minimal position error. The system output is data rich with orientation, velocity and acceleration recorded at 100Hz. Figure 1 illustrates typical use of the sensor technology.

3 Motion Model

The idea presented in this paper involves a multimodal control framework of motions consisting a number of layers of increasing abstraction. The view is held that humans form simpler representations of motion than at the sensory level [Schmidt, 1982]. It is unclear how such a system may be mapped from real motions or in our case motion data. In this work motion was considered along two strains, as a trajectory composed of the joint angle values and at an abstract behaviour level.

3.1 Motion Primitive Layer

Motion primitives are clustered at a trajectory level into a Gaussian Mixture Model (GMM) using the Expectation Maximization (EM) algorithm to fit the data and Minimum Message Length (MML) criterion to assess the statistical significance of the number of clusters.

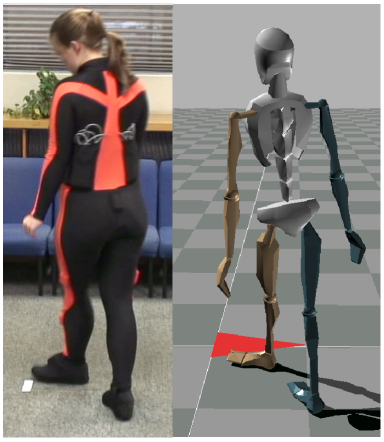

Figure 1: The sensor technology in use.

A variation on AutoClass [Cheeseman and Stutz, 1996] is used to cluster the angle data into a plausible number of Gaussian mixtures. The data is normalised and the accuracy of measurement (aom), which quantizes each continuous variable for clustering, is given as a percentage of the data variance. The use of MML is a quantitative embodiment of 'Occam's Razor' where model selection is a trade-off between simplicity and 'goodness of fit'. First, the algorithm postulates a model, fits the data using EM and then evaluates it based on the estimated code length to express the model and the accuracy of the model.

From information theory, the minimum coding length of any message is given by

$$L(E) = -\log(P(E)). \tag{1}$$

where E is the evidence or data and from Bayes' theorem

$$P(E|H) \cdot P(H) = P(H \cap E). \tag{2}$$

where H is a probable hypothesis. Maximising the probability that the evidence supports the hypothesis is equivalent to minimising the message length.

$$\arg\max P(H \cap E) \equiv \arg\min(-\log(P(E|H) - \log(P(H))). \tag{3}$$

If the resultant model is sufficiently complex a close approximation to the original trajectory can be extracted but due to the high dimensions over-fitting can still occur.

3.2 Abstraction Layer

At a higher management or abstract level particular behaviours or manoeuvres may be comprised of an ordered

set of motion primitives, which can reconstruct a full trajectory. Extraction of these behaviours is also handled as a segmentation problem, which could be approached in a variety of ways presented in the next section.

With the abstract segmentations and motion primitives considered in parallel, behaviours can be defined by a set of motion primitives. If similar sets are observed they correspond to similar behaviours and therefore the regions of data bounded by segmentations can be merged. If any behaviour sets share a membership proportion of motion primitives above a given threshold they become candidates for merging. However, some behaviours overlap and appear to expand the true coverage without incorporating a membership degree based upon the number of appearances. Motion primitives with low membership may be discarded for one sets if it has high membership for a neighbouring set, as can happen in repetitive sequences. This method only applies when repeated patterns emerge from the sequence.

4 Motion Segmentations

The goal of the motion segmentation is to separate grossly different behaviours by identifying significant events within the time series sequences. Events that may indicate abrupt or more gradual changes in the data structure. The procedures attempted were a variety of simple reduction techniques with an emphasis on producing symbolic representations of the data.

4.1 Activity level

The simplest method was to use the accelerations to detect changes in activity in the motion. The magnitude of the accelerations a_i were scaled by the corresponding mass m_n for $n = 1, 2, ...k$ where k is the number of segments and added across D dimensions.

$$A = \sum_{i}^{D} m_n |a_i|. \tag{4}$$

This results in one time series signal that can be reduced into symbolic form using Symbolic Aggregate Approximation (SAX) [Lin and Keogh, 2007] and analysed as a string of symbols to detect significant trend changes. The SAX technique transforms a unidimensional time series into symbols by introducing piecewise aggregate approximation (PAA) and magnitude breakpoints to quantize the signal. Two parameters required for this method are the symbol alphabet size α or number of breakpoints and the size of the piecewise segments. An example symbolized activity trend is shown in Figure 2.

With a string of symbols describing the trend, analysis is greatly simplified. The method proposed here for separating behaviour involved a distance calculation between two adjacent sets of symbols. If this distance exceeded

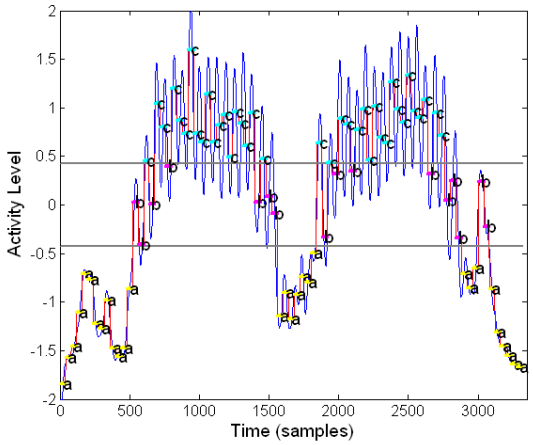

Figure 2: Example of activity level quantized into symbolic form.

a pre-defined threshold the two strings belong to different behaviours. With a set size of w_s each inter-symbol distance between sets was calculated by the time difference multiplied by the alphabet difference. The measure effectively gives preference to high gradients in the original signal, which were retained through the aggregation and symbol steps.

$$d = \sum_{i}^{N} \sum_{j=i-w_s}^{i-1} \sum_{k=i}^{i+w_s-1} |S_j - S_k|((2w_s - 1) - (k - j)). \tag{5}$$

where N is the length of the symbol sequence and S is a numbered alphabet index. Segmentations are taken at the peaks of the distance curve when the difference between a previous minimum and running maximum is greater than a threshold R as indicated in Figure 3.

The method can be tuned to detect smaller or less pronounced behaviours by varying parameters such as the alphabet size and sampling rate of the symbols. The threshold R or set size w_s can also influence the number of segments detected.

4.2 Zero Crossings

Another set of significant events may be the zero crossing of a velocity signal (in changing direction) or acceleration signal (in shifting inertia). The significance of Zero Velocity Crossings (ZVC) in motion segmentation was proposed by [Fod *et al.*, 2002], where two ZVCs within 300ms produced a segmentation as a preprocessing step before applying PCA to each finite time series.

Although these crossings tend to be sparse throughout a motion, it is hypothesised here that they would be concentrated around the start or completion of a motion or during periods of reduced activity. Since they are also unlikely to occur simultaneously across each DOF, a

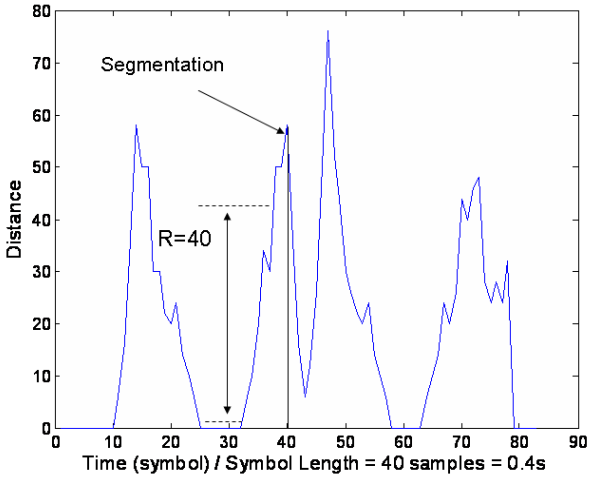

Figure 3: Distance calculation across time for adjacent symbolic sets and segmentation process.

count of zero crossings was implemented across all DOF within a neighborhood of samples w_{zc}. The result is a unidimensional time series that is analysed using the same symbolic methods as shown in the previous section.

Once again the parameters can be varied to similar effect as the previous method with the inclusion of the neighborhood size value w_{zc} for which sensible values lie is the range $5 \leq w_{zc} \leq 30$ and produce consistent segmentations.

4.3 Probabilistic PCA segmentation

In the next approach Probabilistic PCA (PPCA) was used to compare behaviours as distributions. This technique was used on quaternion joint angles in segmenting motion capture databases in [Barbic *et al.*, 2004]. Our work investigates segmentations based on acceleration and Euler angle data.

The first N samples of the data set are modelled by a Gaussian distribution with a mean $\bar{x}$ and covariance C defined by

$$C = \frac{1}{N-1} V \tilde{\Sigma}^2 V^T. \quad (6)$$

The vector V is the set of principle component variances and the square matrix Σ has nonnegative decreasing singular values on its diagonal obtained from singular value decomposition (SVD). In PCA a proportion of the eigenvalues are discarded to set the minimum variance coverage of the new PCs. $\tilde{\Sigma}$ is produced from Σ by replacing these discarded values with noise.

Another set of the data from $N+1$ to $N+T$ is compared to the Gaussian distribution defined by $\bar{x}$ and C using the Mahalonobis distance,

$$H = \frac{1}{T} \sum_{i=N+1}^{N+T} (x_i - \bar{x})^T C^{-1} (x_i - \bar{x}). \quad (7)$$

Table 1: Percentage match of algorithm segments to manual segments for different input data.

Method	Acceleration	Velocity	Angle
ZC	0.75	0.73	0.82
Activity	0.86	0.83	0.89
PPCA	0.90	0.82	0.84

The value of N is updated incrementally by a constant number of frames, Δ, and the calculation performed again. Initially the distance is high, reduces significantly and remains low while the behaviours are the same, it increases to a peak when the new data is not modelled by the distribution. Segmentations are taken at the peak and the process started again at this new reference point.

5 Results

The techniques described in previous sections were tested on a range of motion data. Each segmentation result was evaluated in comparison to results generated manually. The some experimentation was required to select appropriate parameters and visualisations were used of the resulting segmentations to verify the separation of distinct behaviour.

5.1 Data

The data sets used in the analysis include several sequences of walking patterns, running, sporting action, and stretches. There are over 30,000 frames or 5 minutes of recordings at 100Hz. The first set is a collection of 4 walking sequences including changes of direction. Another set contains running, jogging and the transitions between. The next is a collection of football kicks and the final set is a mixture of stretches, walking, climbing in one sequence.

5.2 Evaluation

Behaviour segmentations were evaluated against a human perception of distinct behaviours. A group of individuals independently examined the motion sequences and recorded a set of segmentation points. The range of the most common segmentation points were compared with each algorithm result. Each algorithm was rated based on their match with the manual segments to within an error of 100 samples.

Although each method can produce vastly different results by tuning the parameters, a set of values was chosen for each technique which appeared to perform best for all the training data. The percentage of algorithm segment points that aligned with manual points are shown in Table 1. This percentage is the average performance over all of the data, however some algorithms performed better on cyclic motions than others.

Parameters

Varying the parameters in each method can affect the detection of behaviours. The activity level and zero crossing methods share all parameters including the symbol frequency f, alphabet size α, length of string comparison w_s distance threshold R, except for the crossing neighbourhood w_{zc}. Through experimentation these parameters were chosen as $f = 20, \alpha = 8, w_s = 4, R = 40$, and for zero crossings $w_{zc} = 20$ corresponding to 200ms. Decreasing f or R and increasing α introduces greater sensitivity. Increasing this sensitivity tends to result in more oscillatory symbols requiring a different symbolic analysis than applied in this paper.

The parameters for the PPCA segmentation were $T = 80, \Delta = 5, R = 20$ with $N = T$ initially. Reducing T allows the detection of more rapid behaviours while the threshold R influences the sensitivity of segmentation. For segmenting angle data R was increased to 100 to reduce the segmentations found.

Visualisation

The results can also be analysed visually by observing the membership states of the behaviours. Figure 4 displays a visualisation of the state model where the lower level motion primitives (blue) are drawn in a connected graph or finite state model (FSM). The behaviours (red) are defined by their motion primitive members. Each motion primitive pose can be observed using the mean of the Gaussian mixtures.

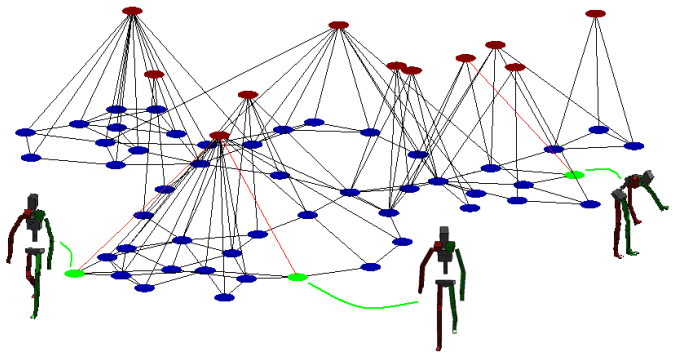

Figure 4: Motion primitive (blue) in a FSM and the segmented behaviours (red) connected to their primitive members. Illustrated states are highlighted green.

A behaviour may also be visualised by superimposing all of the motion primitives on the same plot as in Figure 5. The set of motion primitives clearly belong to different behaviours with Figure 5(a) displaying a bending and reaching motion while (b) is a walking motion.

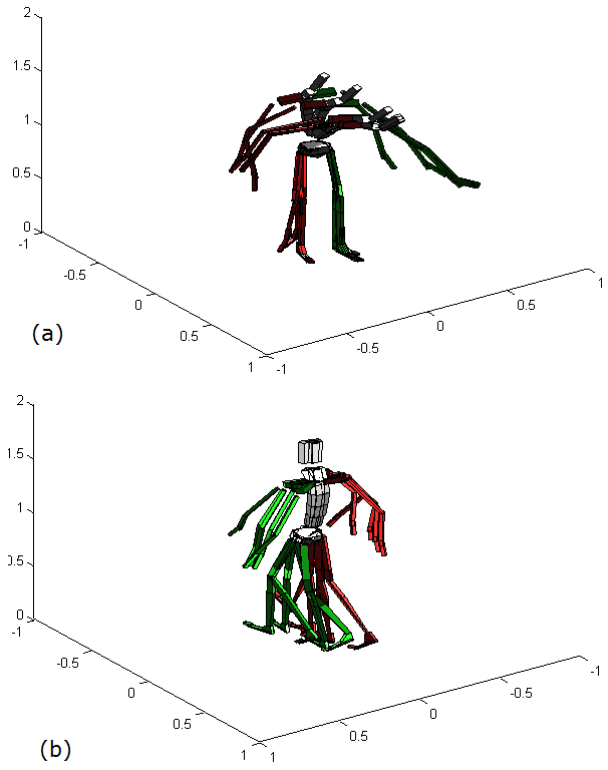

Figure 5: Separation of different behaviours by superimposing primitives.

6 Conclusions

This paper presented a technique for modelling human motion for imitation robotics which was divided into two levels, the sequence of trajectory motion primitives and the abstract organisation level. Motion primitives were extracted from observations via a GMM algorithm optimised by MML encoding. Segmentations at an abstract level were implemented and compared on a range of data sets.

Among the methods used PPCA performed well on acceleration data but often consumed more processing time. Despite significant simplifications of the data in the other methods their performance was still close to manual separation across the data with carefully selected parameters.

There is scope for further experimentation in this area once multidimensional signals are converted to symbolic form due to the wealth of techniques available in text data mining. More sophisticated symbolic sequence algorithms need to be used when there are changing oscillatory patterns.

The modelling procedure segments motion into sequence of subtasks. For humanoids, it is potentially useful to reorganise these elements to perform new tasks by

separating generalisations into hierarchical levels. Future work will extend these techniques into robot imitation and control.

References

[Barbic *et al.*, 2004] Barbic, J., Safonova, A., Pan, J., Faloutsos, C., Hodgins, J. K., and Pollard, N. S. Segmenting motion capture data into distinct behaviors. *In Proceedings of Graphics interface*, p. 185-194, London, Ontario, Canada, May 17-19, 2004.

[Beaudoin *et al.*, 2008] P. Beaudoin, M. van de Panne, P. Poulin, and S. Coros. Motion-Motif Graphs. *Proc. Symposium on Computer Animation 2008*, To appear, July 2008.

[Calinon *et al.*, 2007] S. Calinon, F. Guenter, and A. Billard. On learning, representing and generalizing a task in a humanoid robot. *IEEE Transactions on Systems, Man and Cybernetics: Part B*, Vol. 37, No. 2, pp. 286-298, 2007.

[Cheeseman and Stutz, 1996] P. Cheeseman, J. Stutz. Bayesian Classification (AutoClass): Theory and Results. In *Advances in Knowledge Discovery and Data Mining*, U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, Eds. American Association for Artificial Intelligence, Menlo Park, CA, 153-180.

[Fod *et al.*, 2002] Fod, A., Mataric, M. J., and Jenkins, O. C. Automated Derivation of Primitives for Movement Classification. *Autonomous Robots*, 12, 1 (Jan. 2002), 39-54. DOI= http://dx.doi.org/10.1023/A:1013254724861.

[Inamura *et al.*, 2004] Inamura, T., Toshima, I., Tanie, H. and Nakamura, Y. Embodied symbol emergence based on mimesis theory. *International Journal of Robotics Research*, 23(4-5): pp. 363-377.

[Kulic *et al.*, 2008] D. Kulic, J. W. Takano, and Y. Nakamura. Incremental Learning, Clustering and Hierarchy Formation of Whole Body Motion Patterns using Adaptive Hidden Markov Chains. *International Journal of Robotics Research*, Vol. 27, No. 7, pp. 761-784, 2008.

[Lin and Keogh, 2007] J. Lin, E. Keogh, L. Wei, S. Lonardi. Experiencing SAX: a Novel Symbolic Representation of time series. *Data Mining and Knowledge Discovery Journal*, Vol. 15, No. 2, pp. 107-144, 2007.

[Schmidt, 1982] Schmidt, R. A. Motor control and learning: A behavioural emphasis. Champaign, IL: Human Kinetics Publishers.

[Shon *et al.*, 2006] A. P. Shon, K. Grochow, A. Hertzmann and R. P. N. Rao. Learning shared latent structure for image synthesis and robotic imitation. *Advances in Neural Information Processing Systems*, in Y. Weiss, B. Schlkopf and J. C. Platt (eds), MIT Press, Cambridge, MA.

[Sun *et al.*, 2006] Chao Sun, David Stirling, and Fazel Naghdy. Human Behaviour Recognition with Segmented Inertial Data. In *ARAA Australasian Conference on Robotics and Automation*, 1–9.

[Tanaka *et al.*, 2005] Tanaka, Y., Iwamoto, K., and Uehara, K. Discovery of Time-Series Motif from Multi-Dimensional Data Based on MDL Principle. *Machine Learning*, 23(4-5): pp. 363-377.

[Wang *et al.*, 2008] Wang, J.M., Fleet, D.J., Hertzmann, A. Gaussian Process Dynamical Models for Human Motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.30, no.2, pp.283-298, Feb. 2008.