ESANN'2007 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 25-27 April 2007, d-side publi., ISBN 2-930307-07-2.

Human Motion Recognition using Nonlinear Transient Computation

Nigel Crook and Wee Jin Goh

School of Technology - Department of Computing
Oxford Brookes University, Wheatley Campus, Oxford - United Kingdom

Abstract. A novel approach to human motion recognition is proposed that is based on a variation of the Nonlinear Transient Computation Machine (NTCM). The motion data used to train the NTCM comes from point-light display video sequences of a human walking. The NTCM is trained to distinguish between sequences of video frames that depict coordinated walking motion from those that depict uncoordinated (random) motion.

1 Introduction

Artificial neural networks (ANNs) have been used extensively in recent years for the recognition of human motion [1, 2]. ANN approaches to human motion recognition broadly divide into two categories: those that use what could be described as standard ANN technologies (e.g. layered feedforward networks trained with variants of Error Backpropagation [3]), and those that use more biologically plausible neural network models [2]. The approach presented in this paper does not fall into either of these categories. Instead, a novel method for motion recognition is proposed based on Nonlinear Transient Computation [4, 5].

Nonlinear Transient Computation provides a means of classifying time-varying input signals using the dynamics generated by a nonlinear attractor. The input signals momentarily perturb the dynamics of the nonlinear system away from the attractor. This perturbation is proportional to the input that caused it. The subsequent evolution of the system follows a transient back to the attractor forming a trajectory that is uniquely determined by the original input signal. A simple linear readout neuron can then be trained to recognise the transients that are generated by distinct classes of input signals.

The properties of the Nonlinear Transient Computation machine (NTCM) have been studied in some detail elsewhere [4, 5, 6]. This paper is the first to explore the possibility of applying the NTCM to the problem of recognising human motion within video sequences of point-light displays. In this study, a variation of the NTCM (denoted as the LTCM) is used that employs the Lorenz equations [7] to generate the chaotic attractor [6]. The LTCM is trained to distinguish between sequences of video frames of coordinated human (walking) motion from sequences of video frames showing uncoordinated motion (i.e. sequences of randomly selected frames from a human walking motion video).

2 The LTCM model

The LTCM consist of two components. The first is the device described as N_T that provides the nonlinear attractor that gives rise to the transients induced by external input. The second is the set of simple linear readout devices N_R, each trained to recognise transients that correspond to a class of input signals.

The device denoted as N_T can be any system that provides a nonlinear attractor. For example, it could be a large pool of recurrently connected neurons configured to generate near-chaotic firing patters [8]. Our previous work with the NTCM has used the Nonlinear Dynamic State neuron to provide the attractor for N_T [9]. In this paper, N_T is modelled using the chaotic attractor provided by the Lorenz system, which has been thoroughly studied within the field of dynamical systems theory [7]. The attractor produced by the Lorenz equations is a two-dimensional surface in three dimensional phase space. Consequently, this system can be perturbed by an input vector with up to three dimensions that can be arbitrarily assigned to its state variables x, y and z (see [6] for dealing with input vectors with more than three dimension):

$$\begin{aligned}
\dot{x_i} &= \sigma(y_i - x_i) + w_x I_x(t) \\
\dot{y_i} &= x_i(\tau - z_i) - y_i + w_y I_y(t) \\
\dot{z_i} &= x_i y_i - \beta z_i + w_z I_z(t)
\end{aligned}$$

where w_k is the weight of the kth dimension of input vector I which perturbs the system at time t, and x_i, y_i and z_i are the state variables of the ith trajectory of N_T (see below). The chaotic nature of this system guarantees that two similar but distinct perturbations will eventually result in very different trajectories around the attractor (a property often described as *sensitivity to initial conditions*). However, even with a system that is as chaotic as the Lorenz attractor, similar perturbations will cause similar trajectories *in the short term*. In other words, similar inputs to N_T will result in transients that are initially similar but that later diverge as the two transients evolve. This can been exploited to adjust the sensitivity of the NTCM to the presence of noise in the input signals [5, 4].

Given that the perturbation caused by an input is directly proportional to that input, it becomes possible to construct a classifier based on the transients coming from the attractor. An observer trained on the transients will be able to use the afore mentioned property to differentiate between classes of input. Due to the chaotic nature of the system, transients will diverge as the system evolves along time. It is this nonlinear divergence that allows for the classification of patterns that are otherwise nonlinearly separable. There isn't scope within this paper to fully explore the function of the non-linear attractor provided by N_T in the pattern recognition process; this has been dealt with elsewhere [6]. However, it is worth briefly outlining that the LTCM has been successfully applied to the classification of static input, such as the IRIS dataset. The results obtained are comparable with other well known methods of pattern classification [6].

ESANN'2007 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 25-27 April 2007, d-side publi., ISBN 2-930307-07-2.

3 Motion Recognition

This paper proposes a novel approach to human motion recognition based on the LTCM. In this study, short video sequences of point-light displays are given as input to the LTCM, which then has to determine whether or not each sequence is showing a human walking motion. Each frame in the point-light display sequences used in this study has up to 13 points that are located on the major joints of the human body (Figure 4). As the sequence progresses certain points disappear through the occlusions caused by limbs moving over an area of the body. The approach to motion recognition proposed here does not require prior knowledge of which points correspond to which joints in the body. Neither does it require the tracking of points from one frame to another. Consequently, the appearance or disappearance of points from the frames as the sequence progresses does not cause any difficulties for the LTCM motion recognition system.

In the system presented here, the transient device N_T maintains 13 numbered trajectories simultaneously in the same attractor space. Each point within a point-light display frame causes a perturbation on one of these 13 trajectories, with a one-to-one correspondence from points to trajectories for each frame. Points are assigned to trajectories in an arbitrary but consistent manner: For each frame the points are sorted using the y-coordinate as the primary sort key and the x-coordinate as the secondary sort key. The sorted points are then assigned to trajectories in increasing order of trajectory number. If there are $n < 13$ points within a given frame, then the trajectories numbered $n + 1$ to 13 remain unassigned for that frame. At the start of the frame sequence all 13 trajectories are initialised to an arbitrarily chosen point within the basin of attraction of N_T.

The transient device N_T is centrally located with respect to its receptive field on the point-light display. As each frame of the motion is displayed, the points in the frame are assigned to the 13 trajectories as described above. The relative position of a point from the center of the receptive field forms a 2D vector which is then scaled and added to the assigned trajectory in the x-y plane of the phase space (in this application, the trajectories are not perturbed in the z dimension). Thus each trajectory that has been assigned a point from the frame is perturbed in a direction that is proportional to the vector produced by the assigned point. The 13 trajectories are evolved for a fixed number of times steps (5 in the experiments reported below) between the presentation of the frames. Once all the frames have been presented to the LTCM, the 13 trajectories are evolved for a further fixed number of time steps. The time series for the three state variables are then summed to form a single time series I_i for input pattern i, which is then sampled a regular intervals of k time steps ($k = 10$ in the experiments reported below) to form the input R_i for the readout device N_R:

$$\begin{aligned} I_i(t) &= x_i(t) + y_i(t) + z_i(t) \\ R_i &= \{I_i(t) : t \bmod k = 0\} \end{aligned}$$

In the experiments reported below, the readout device N_R is a single perceptron trained using the generalised delta rule.

The perturbations to each of the 13 trajectories push them away from the attractor in a direction determined by the input vectors, causing each trajectory to enter a transient back towards the attractor. These transients are proportional to the sequence of perturbations that caused them, and hence are proportional to the point-light sequences that were given as input to the LTCM.

4 Experimental Results

A 75-frame point-light display video sequence of a walking human provided the data for the experiments reported here. Positive examples of walking motion were generated by taking sequences consisting of 10 consecutive frames from the walking motion video (See Figure 1). A total of 65 10-framed examples of walking sequences were extracted from the video. Negative examples of walking motion were constructed by taking 65 sets of 10 randomly chosen frames from the walking video. Hence, although each frame of a negative sequence is a valid walking pose, the whole 10-frame random sequence does not constitute a recognisable walking motion. In a sense, the positive examples of walking are instances of coordinated motion, whilst the negative examples are instances of uncoordinated (random) motion.

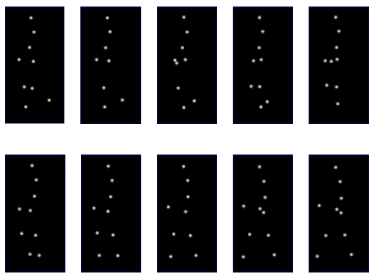

Fig. 1: An example of a 10-frame walking sequence

The readout device of the LTCM is a simple perceptron capable only of correctly partitioning linearly separable clusters of input patterns. Hence the chaotic attractor provided by N_T is a crucial element of the LTCM's motion recognition apparatus enabling it to solve linearly inseparable problems. To show that the walking motion recognition problem is linearly inseparable, and to demonstrate the utility of the chaotic attractor in solving this problem, the walking motion data will also be applied to a single perceptron (i.e. the LTCM without the chaotic attractor given by N_T). The motion data is presented to the perceptron as follows: each frame of the sequence is *flattened out*, so that the coordinates of the points p in frame f are presented consecutively as follows :

	Training Set			Test Set				
		Accuracy			*Accuracy*		*Thresh(0.8)*	
Classifier	*MSE*	*95%*	*80%*	*MSE*	*95%*	*80%*	*Sen.*	*Spec.*
P:	0.135	83.8%	86.2%	0.438	48.5%	52.3%	43.7%	74.5%
LTCM:								
A(0..50)	0.101	67.1%	80.9%	0.212	57.7%	67.7%	72.4%	87.4%
B(50..200)	0.002	99.8%	99.8%	0.141	66.9%	73.8%	75.4%	80.9%
C(0..200)	0.003	99.7%	99.7%	0.123	70.0%	80.0%	83.8%	83.6%

Table 1: Results from the 10-fold cross validation experiments.

$[p1_x^f, p1_y^f, p2_x^f, p2_y^f, ...p13_x^f, p13_y^f]$; All 10 frames of a sequence are presented as one input pattern (hence the perceptron has 260 inputs). The x and y coordinates of each point are scaled to values in the range [0..1]. Note that this encoding of the motion data preserves the ordering of the points in each frame (i.e. point 12 is always the right knee of the walking human, whereas point 13 is always the right ankle). This gives the perceptron information that is not available to the LTCM due to the occlusion of points during the walking sequence. The results from the perceptron will also provide a baseline for these experiments.

10-fold cross validation was used to train and test the LTCM and the single perceptron. The 65 positive and 65 negative examples of walking motion were randomly divided in to 10 sets. The LTCM and the single perceptron were trained on nine of the sets and tested on the tenth set. This was repeated so that all 10 sets were, in turn, used as test data, with the remaining nine sets used as training data in each case.

A key question in this work concerns the extent to which the transients that evolve after the input has been presented contribute to the pattern recognition capabilities of the LTCM. To shed some light on this, three different versions of the LTCM were constructed: in the first (LTCM-A) the readout neuron was only allowed to observe the transients during time steps [0..50] while the input was being presented (each 10 frame sequence is presented to the LTCM in the first 50 time steps of the evolution of the trajectories); in the second version (LTCM-B) the readout neuron only observes the trajectories after the inputs have been presented (i.e. time steps [50..200]); the third version (LTCM-C) is allowed to observe the who trajectory up to time step 200 (i.e. [0..200]).

5 Discussion

The results of the 10-fold cross validation experiments with the perceptron (denoted P) and the three versions of the LTCM (denoted A, B and C) are presented in Table 1. The table shows the mean squared error (MSE) and the accuracy for both the training set and the test set for each experimental run. Two levels of accuracy are shown: the columns headed 95% shows the percentage of outputs from the data set that were within 5% of the target output for each pattern;

similarly the columns headed 80% shows the percentage of outputs that were within 20% of the target outputs. The final two columns show the sensitivity (column headed *Sen.* - the percentage of true postive outputs), and the specificity (column headed *Spec.* - the percentage of true negavite outputs) for all four classifiers on the data set when a threshold of 0.8 is applied to their outputs.

As can be seen from Table 1 the perceptron performs reasonably well on the training data set but poorly on the test set. Given that on average the test sets consist of 50% postive and 50% negative examples of walking motion, the accuracy figures of the perceptron on this data suggest that it performs little better than chance guess on recognising positive examples of walking motion. The poor performance of the perceptron on this data suggests that this walking motion recognition problem is linearly inseperable.

All three versions of the LTCM out-perform the perceptron on this problem. The LTCM-A, which observes the first 50 time steps of the evolution of N_T has the lowest accuracy level of the three LTCMs, but is the best at recognising non-walking motion sequences. LTCM-B, which observes the transient after input (time steps 50-200) performs best in terms of recognising true positives with the threshold output. However, LTCM-C, which observes the whole transient of N_T has the highest level of accuracy, whilst maintaining moderately high sensitivity and specificity.

Whilst these results are not conclusive, they suggest the attractor provided by N_T performs a critical role in the LTCM in enabling it to solve linearly inseparable problems. Furthermore, LTCM has demonstrated an ability to differentiate between coordinated and uncoordinated motion. Future work in this area will compare the NTCM with other approaches to motion recognition.

References

[1] T. Chua. A review of analytical techniques for gait data. part 2: neural network and wavelet methods. *Gait and Posture*, 13:102–120, 2001.

[2] M.A. Giese and T. Poggio. Neural mechanisms for the recognition of biological movements. *Nature Reviews: Neuroscience*, 4:179–192, 2003.

[3] V. Laxmi, J.N. Carter, and R.I. Damper. Biologically-inspired human motion detection. In Michel Verleysen, editor, *10th European Symposium on Artificial Neural Networks*, pages 95–100, Bruges, April 2002. D-Facto, Belgium.

[4] N.T. Crook. Nonlinear transient computation. *Neurocomputing*, in press.

[5] N.T. Crook. Nonlinear transient computation and variable noise tolerance. In M. Verleysen, editor, *Proceedings of 14th European Symposium on Artificial Neural Networks (ESANN'2006)*, pages 509– 514, Bruges, April 2006. d-side, Belgium.

[6] W.J. Goh and N.T. Crook. Pattern recognition using chaotic transients. In *Submitted to 15th European Symposium on Artificial Neural Networks (ESANN'2007)*, 2007.

[7] N.E. Lorenz. Deterministic non-periodic flows. *Journal of Atmospheric Science*, 20(2):130–141, 1963.

[8] Nils Bertschinger and Thomas Natschläger. Real-time computation at the edge of chaos in recurrent neural networks. *Neural Comput.*, 16(7):1413–1436, 2004.

[9] N.T. Crook, W.J. Goh, and M. Hawarat. The nonlinear dynamic state neuron. In M. Verleysen, editor, *Proceedings of 13th European Symposium on Artificial Neural Networks (ESANN'2005)*, Bruges, April 2005. d-side, Belgium.