Novelty Detection in Video Surveillance Using Hierarchical Neural Networks

Jonathan Owens[1], Andrew Hunter[2] and Eric Fletcher[1]

[1] Centre for Adaptive Systems, University of Sunderland, UK
{jonathan.owens,eric.fletcher}@sunderland.ac.uk
[2] Department of Computer Science, University of Durham, UK
andrew1.hunter@durham.ac.uk

Abstract. A hierarchical self-organising neural network is described for the detection of unusual pedestrian behaviour in video-based surveillance systems. The system is trained on a normal data set, with no prior information about the scene under surveillance, thereby requiring minimal user input. Nodes use a trace activation rule and feedforward connections, modified so that higher layer nodes are sensitive to trajectory segments traced across the previous layer. Top layer nodes have binary lateral connections and corresponding "novelty accumulator" nodes. Lateral connections are set between co-occurring nodes, generating a signal to prevent accumulation of the novelty measure along normal sequences. In abnormal sequences the novelty accumulator nodes are allowed to increase their activity, generating an alarm state.

1 Introduction

To help CCTV surveillance operators maintain adequate attention levels, the next generation of automated visual surveillance systems will be attention-focussing filters that restrict events presented to the operator to those that fall outside of some definition of normality [1]. The novelty of an observed trajectory can be measured by building probabilistic descriptions of the relationship between trajectory segments [2], [3]. Neural networks have been used to create distributions of full length trajectories, or partial trajectories of significant length, where trajectory segments are weighted relative to their temporal position in the sequence [4], [5].

The system described in this paper uses a hierarchical network of self-organising layers to represent trajectory segments, and a top layer of nodes with binary lateral connections and novelty accumulator nodes to measure the novelty of sequences of trajectory segments. Rather than requiring the completed trajectory to be submitted to the network, novelty is measured as trajectories are traced across the input layer, so an alarm can be generated while the event is occurring.

The neural network is one component of a hybrid system, which uses a temporal low-pass filter to construct a background image, and segments moving objects by background differencing. Existing objects are assigned to the segmented connected components by a minimum-distance match across feature vectors. Short-term novelty such as local erratic motion is detected by a self-organising map [6], which operates alongside the network described in this paper, which detects long-term novelty.

2 Hierarchical Self-Organising Network Structure

Self-organisation in a hierarchical structure is the basis for many powerful models, from the strictly computational [7], [8], to those that model specific biological networks [9], [10]. Our basic architecture is shown in figure 1.

Each node uses a trace activation rule, which was first hypothesised as a means by which the higher visual cortices may learn object invariance [11], and was used to that effect in [10] for learning invariant recognition of objects subject to spatial transformations. The feedforward nodes have overlapping receptive fields, except in layer 1, where the nodes simply partition the 2D image plane into fixed width cells. The final layer nodes have modifiable binary lateral connections and novelty accumulator nodes.
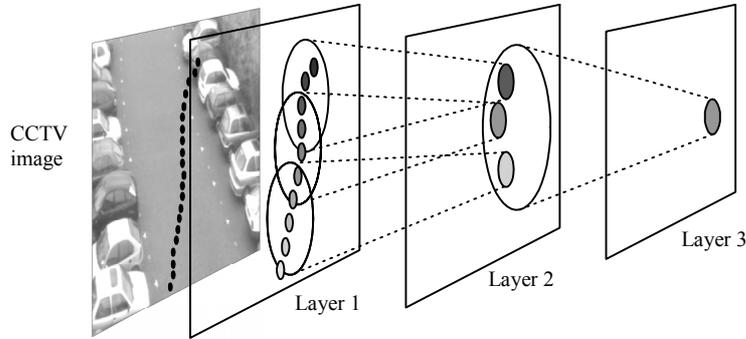

Fig. 1. General architecture, with trace activation illustrated as a decrease in greyscale value of nodes participating in the trace (for clarity, only a subset of nodes are shown)

2.1 Learning

Training is partitioned such that modification of feedforward connections to layer k is completed before training begins at layer $k+1$. Subsequently, lateral weights between top layer nodes are modified to reflect normal co-occurrences of top layer nodes.

The object tracker passes the centroid of the tracked object to the first layer of the hierarchical network. The activity of layer 1 nodes signals the recent presence of the tracked object within their receptive fields by the following activation rule:

$$x_i(t) = \begin{cases} 1 & [r(t),c(t)] \in R_i \\ x_i(t-1)*\gamma & \text{otherwise}, \quad (\gamma < 1) \end{cases} \tag{1}$$

Where x_i is the activity of first layer node i, γ is the trace rule decay constant, $[r(t),c(t)]$ is the centroid of the tracked object at time t and R_i is the receptive field. Activity of nodes in higher layers is calculated as follows:

$$a_i(t) = \sum_{j \in R_i} w_{ij}(t)x_j(t) \tag{2}$$

$$x_i(t) = \begin{cases} 1 & a_i(t) = \max_n(a_n(t)) \\ x_i(t-1) * \gamma \text{ otherwise} & (\gamma < 1) \end{cases} \quad (3)$$

Where a_i is the weighted sum of the inputs to node i, R_i is the receptive field of node i, x_j is the activity of input node j, w_{ij} is the synaptic weight between node j and i, x_i is the trace activity of node i and γ is the decay constant. The winning node has its activity set to 1, while all other node activities decay according to the trace rule (1). The localised receptive fields and trace activation allow nodes to integrate inputs across a small spatio-temporal window.

Learning only occurs at the winning node at each iteration, and follows the general Hebb rule for synaptic modification [11]. In general,

$$\Delta w_{ij} = \alpha x_j x_i \big|_{j \in R_i} \quad (4)$$

Where x_i is the post-synaptic activity, x_j is the presynaptic activity and α is a learning rate. The sum of the weights impinging on a post-synaptic node is normalised to 1. Learning "chunks" normal trajectory segments into progressively larger segments at higher layers. Synaptic modification is completed in layer 2, before learning begins in layer 3.

Each layer 3 feedforward node has a corresponding "novelty accumulator" (NA) node, which is silent during learning, and is used in novelty detection. Each layer 3 node has binary lateral connections to all NA nodes in the layer, as shown in figure 2.
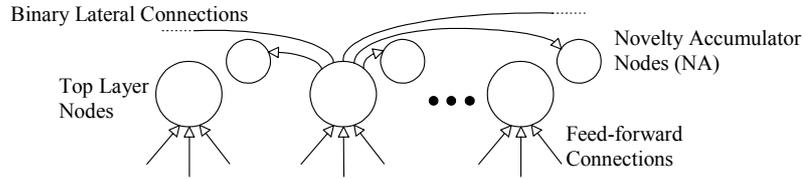

Fig. 2. Inhibition and lateral excitatory connections at the top layer of the network

The lateral binary connections are modified when feedforward learning has been completed. With lateral weights set initially to zero, the training set is submitted to the feedforward section of the network, which produces sequences of winning nodes at the top layer. Binary lateral connections are set, subject to an ordered co-occurrence rule, shown below:

$$l_{ij}(t) = \begin{cases} 1 & [\tau_i > \tau_j] \\ l_{ij}(t-1) & \text{otherwise} \end{cases} \quad (5)$$

Where l_{ij} is the binary weight between pre-synaptic node j and post-synaptic node i, τ_j and τ_i are the onset times of pre- and post-synaptic activity. A connection is set if the onset time of the post-synaptic node activity occurs after the onset of pre-synaptic node activity, during a normal trajectory. Therefore, the binary weights will be set between any given node and any other nodes that have followed it in any trajectory in the training set. The binary connections give a simple indication of the co-occurrence of top layer nodes subject to an ordering constraint.

2.2 Novelty Detection

To classify the observed trajectory, the NA nodes at the top layer are switched on and activity traces are propagated through the network. The winning node at the top layer has its activity set to 1, and the non-winning node and NA node activities are calculated as follows:

$$a_k = a_i l_{ki}\big|_{k \neq i} \qquad (6)$$

$$n_k(t) = \begin{cases} n_k(t-1) + \Delta n & [a_k = 0] \\ n_k(t-1) & [a_k \neq 0] \end{cases} \qquad (7)$$

Where a_k is the activity of top layer node k, n_k is the corresponding NA node, i is the index of the winning node and Δn is an arbitrary non-zero value. The NA nodes are reset at the beginning of every new trajectory.

The winning node i sends a signal through the binary lateral connections to retard the increase of NA node activity at nodes that are expected to follow in the sequence. An unusual trajectory will produce combinations of nodes dissimilar to sequences in the normal training set, and some NA nodes will not receive a retardation signal through the lateral connections, allowing them to increment their activity (eq. 7). Hence, at each time step, the value of the NA node corresponding to the winning node gives a novelty measure of the current point in the sequence.

It should be noted that the number of top layer nodes active in a sequence is relatively small. Due to the similar widths of the convergent receptive fields at each layer, a node in the top layer can receive a signal from a large area of the input. The progressive "chunking" of trajectory segments that occurs as we progress higher up the network means that, at the top layer, an entire sequence may be represented by as few as 4 or 5 nodes.

3 Training and Results

The network structure and list of learning parameters are shown in table 1. The network was trained with 311 normal trajectories, recorded from a CCTV scene over a period of five days. A human operator examined the data set to remove any unusual trajectories generated by tracking failure or genuinely novel behaviour. A set of 20 normal trajectories and 16 unusual trajectories were obtained for testing.

The classification results are shown in table 2 and a selection of trajectories are shown in figure 3. All unusual trajectories were correctly classified, three examples of which are shown in figs. 3a-c. Figure 3a illustrates the unusual occurrence of a pedestrian leaving one vehicle and entering another, while figs. 3b and 3c are novel due to the co-occurrence of trajectory segments never observed in the same sequence during learning.

Figs. 3d-f show the three false positives. Upon examination of the normal test set, the trajectory in fig. 3d was found to have no similar trajectories among the training data, due to pedestrians typically favouring a more direct route out of the car-park (towards the bottom left of the surveillance image). Figs. 3e and 3f show trajectories

assigned to the normal data set, but which nevertheless show high curvature not found in any other normal examples.

Table 1. Network parameters. Note: Training only occurs at the fan-in connections to layers 2 and 3, layer 3 does not need an activity trace and layer 1 receptive fields are non-overlapping

Layer, n	Layer Size	Receptive field width	Trace decay, γ_n	Learning rate, α_n	Training epochs
1	72 x 54	9 (pixels)	0.8	——	——
2	23 x 32	11 (nodes)	0.8	0.01	100
3	7 x 12	11 (nodes)	——	0.008	100

Table 2. Classification results. Note: The 17 correctly classified normal trajectories are not shown, i.e. trajectories with no novel points

Data	Classification	
	Normal	Unusual
Normal	17	3
Unusual	0	16

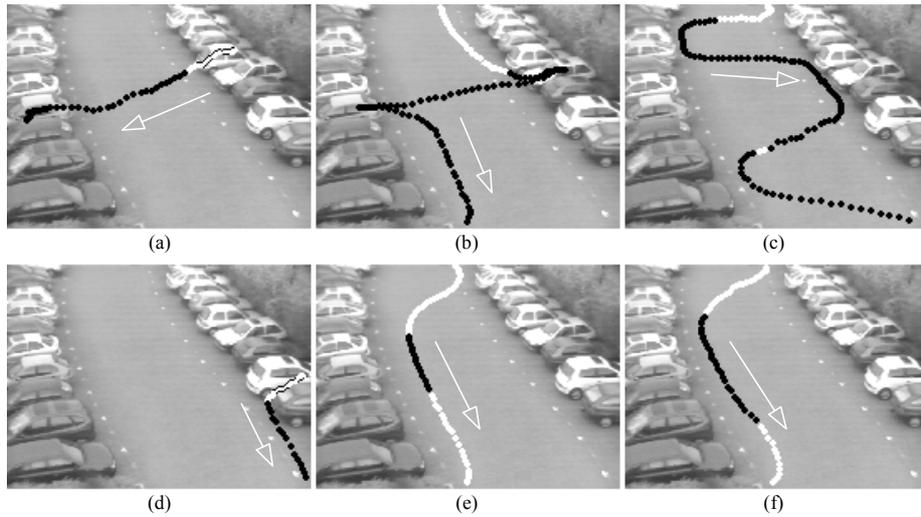(a)   (b)   (c)

(d)   (e)   (f)

Fig. 3. Trajectories with points classified as normal (white) or unusual (black). White arrows indicate direction of motion

4 Discussion

As described in section 2.2, a "retardation" signal through the binary lateral connections prevents the NA nodes which correspond to normal sequences of winning nodes, from accumulating non-zero activity, which is the measure used to classify

trajectories as novel. Due to the converging feedforward receptive fields and relative sparsity of top layer activation, the generality of the "expected" sequences expressed by the binary lateral connections is quite broad. In fact, the system was able to correctly classify a normal test set, previously unseen by the network.

The main advantage of this system compared with other "global" novelty detectors is the continuous nature of the classification, it is not necessary to wait for an entire trajectory, or a significant portion, to be completed before it is submitted for classification. This is important for surveillance applications where crime prevention requires alarm reporting as the event is developing. The NA nodes effectively express the consistency of the current node with preceding nodes in the sequence, with the binary lateral connections giving an indication of an "expected" forward path that may be taken across the top layer.

The misclassified normal trajectories shown in table 2 highlight one of the main problems with self-organising systems, that of sensitivity to the sample distribution of the training data. When such false positives are encountered, the trajectory may simply be added to the normal training set and the system retrained. Future work will address such issues, particularly the matter of incremental learning, so that previously unseen normal behaviour can be learned without needing to retrain the entire network.

References

1. Foresti, G.L., Mähönen, P., Regazzoni, C.S. (eds): Multimedia Video-Based Surveillance Systems: Requirements, Issues and Solutions. Kluwer Academic Publishers
2. Stauffer, C., Grimson, W.E.L.: Learning Patterns of Activity Using Real-Time Tracking. IEEE Trans. PAMI, Vol. 22, No. 8 (2000)
3. Mattone, R., Glaeser, A., Buman, B.: A New Solution Philosophy for Complex Pattern Recognition Problems. In: Foresti, G.L., Mähönen, P., Regazzoni, C.S. (eds): Multimedia Video-Based Surveillance Systems: Requirements. Kluwer Academic Publishers
4. Johnson, N., Hogg, D.: Learning the Distribution of Object Trajectories for Event Recognition. Proc. BMVC, Vol. 2 (1995)
5. Srinivasa, N., Narendra, A.: A Topological and Temporal Correlator Network for Spatiotemporal Pattern Learning, Recognition and Recall. IEEE Trans. Neural Networks, Vol. 10, No. 2 (1999)
6. Owens, J., Hunter, A.: Application of the Self-Organising Map to Trajectory Classification. IEEE Third International Workshop on Visual Surveillance (2000)
7. Fukushima, K., Miyake, S.: Neocognitron: A New Algorithm for Pattern Recognition Tolerant of Deformation and Shifts in Position. Pattern Recognition, Vol. 15, No. 6 (1982)
8. Wang, R.: A Hybrid Learning Network for Shift-Invariant Recognition. Neural Network, Vol. 14 (2001)
9. Liaw, J.-S., King, I.K., Arbib, M.A.: Visual Perception of Translational and Rotational Motion. In: Mohan, R. (ed.): Progress in Neural Networks, Vol. 4: Machine Vision (1997)
10. Stringer, S.M. and Rolls, E.T.: Position Invariant Recognition in the Visual System with Cluttered Environments. Neural Networks, Vol. 13 (2000)
11. Földiák, P.: Learning Invariance form Transformation Sequences. Neural Computation, Vol. 3 (1991)
12. Hebb, D.: The Organisation of Behaviour. Wiley, New York (1949)