Learning of Position-Invariant Object Representation Across Attention Shifts

Muhua Li and James J. Clark

Centre for Intelligent Machines, McGill University,
3480 University Street Room 410, Montreal, Quebec Canada H3A 2A7
{limh, clark}@cim.mcgill.ca

Abstract. Selective attention shift can help neural networks learn invariance. We describe a method that can produce a network with invariance to changes in visual input caused by attention shifts. Training of the network is controlled by signals associated with attention shifting. A temporal perceptual stability constraint is used to drive the output of the network towards remaining constant across temporal sequences of attention shifts. We use a four-layer neural network model to perform the position-invariant extraction of local features and temporal integration of attention-shift invariant presentations of objects. We present results on both simulated data and real images, to demonstrate that our network can acquire position invariance across a sequence of attention shifts.

1 Introduction

A computer visual system with invariant object representation should have consistent neural response, within a certain range, to the same object feature under different conditions, such as at different retinal positions or with different appearance due to viewpoint changes. Most research work done so far has focused on achieving different degrees of invariance based only on the sensory input, while ignoring the important role of visual-related motor signals. However, retrieving visual information solely on the input images can cause ill-posed problems. And the extraction of invariant features from visual images which contain overwhelming information is usually slow.

From the viewpoint of active computer vision, an effective visual system can selectively obtain useful visual information with the cooperation of motor actions. In our opinion, visual-related self-action signals are crucial in learning spatial invariance, as they provide information as to the nature of changes in the visual input. Especially, selective attention shifts could play an important role in the visual systems as to focus on a small fraction of the total input visual information ([6], [9]), to perform visual related tasks such as pattern and object recognition, as implemented in several computational models ([1], [8], [12], and [13]). Shifting of attention enables the visual system to actively and efficiently acquire useful information from the external environment for further processing. Therefore, it is conceivable to hypothesize that the learning of invariant presentation of an object might not need complete visual information about the object, instead it can be learned from those attended local features of the object across a sequence of attention shifts.

L. Paletta et al. (Eds.): WAPCV 2004, LNCS 3368, pp. 57–70, 2005.

Here we will focus on the learning of position invariance across attention shifts, with position-related retinal distortions taken into consideration. The need for developing position-invariance arises due to projective distortions and the non-uniform distribution of visual sensors on the retinal surface. These factors result in qualitatively different signals when object features are projected onto different positions of the retina. In order to produce position invariant recognition the visual system must be presented with images of an object at different locations on the retina. Position invariance can be learned by imposing temporal continuity on the response of a network to temporal sequences of patterns undergoing transformation ([2], [4], [5], and [10]). However, when the motion of the objects in the external world produces the required presentation of the object image across the retina, the difference in the appearance of a moving object is generally greater as the displacement increases. This may cause problems in some of the current position-invariant approaches. For example, [4] reported that the learning result was very sensitive to the time scale and the temporal structure in the input.

We introduce attention shifts as a key factor in the learning of position invariance. For the task of learning position invariance, the advantage of treating image feature displacements as being due to attention shifts is the fact that attention shifts are rapid, and that there is a neural command signal associated with them. The rapidity of the shift means that learning can be concentrated to take place only in the short time interval around the occurrence of the shift. This focusing of the learning solves the problems with time-varying scenery that plagued previous methods, such as those proposed by Einhäuser et al. [4] and Földiák [5].

In this paper we present a temporal learning scheme where knowledge of the attention shift command is used to gate the learning process. A temporal perceptual stability constraint is used to drive the output of the network towards remaining constant across temporal sequences of saccade motions and attention shifts. We implement a four-layer neural network model, and test it on both real images and simulated data consisting of various geometrical shapes undergoing transformations.

2 Learning Model

We refer to *local feature* as the visual information falling within the attention window, which includes either a whole object at low resolutions or parts of an object in fine details. Therefore, the learning of position invariance is achieved at two different levels: one is at a coarse level where position-invariant representation of local features is learned; and the other is at a fine level, where the position-invariant representation of an object as a whole with high resolution is learned by temporally correlating local features across attention shifts.

The overall model being proposed is composed of two sub-modules, as illustrated in Figure 1. One is the *attention control module*, which generates attention-shift signals according to a dynamically changing saliency map mechanism. The attention-shift signals are used to determine the timing for learning. The module obtains as input local feature images from the input raw retinal images via a position-changing attention window associated with an attention shift. The second sub-module is the *learning module*, which performs the learning of invariant neural representations across attention shifts

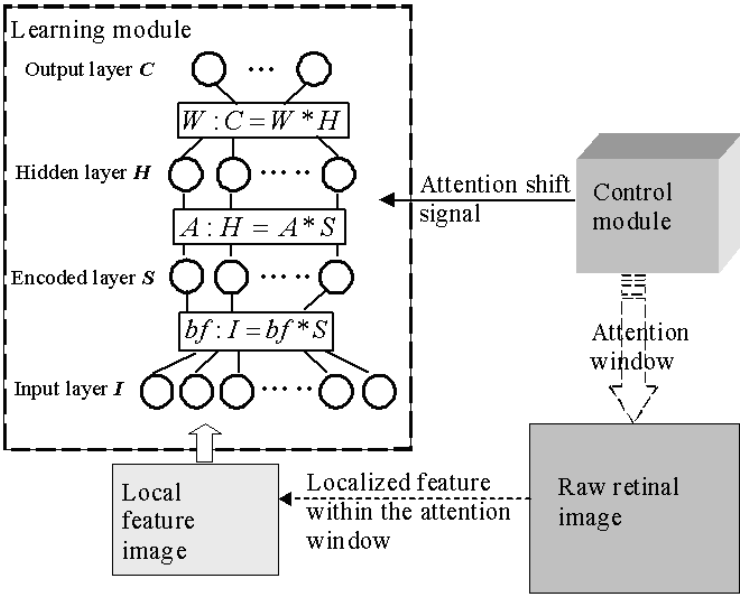

Fig. 1. System structure of the proposed neural network model

in temporal sequences. Two forms of learning, position invariant extraction of local features, and attention-shift invariant integration of object representation (an object is a composition of a set of local features), are triggered by the attention-shift signals. A temporal perceptual stability constraint is used to drive the output of the network towards remaining constant across the sequence of attention shifts.

2.1 Local Feature Selection Across Attention Shifts

Attention shift information is provided in our model by the *attention control module*. This module receives the retinal image as input. It constructs a saliency map mechanism [7] which is used to select the most salient area as the next attention-shift target. Currently our implementation uses grey-level images and we use orientation contrast and intensity contrast as saliency map features to form the saliency map. Intensity features, $I(\sigma)$, are obtained from an 8-level Gaussian pyramid computed from the raw input intensity, where the scale factor ranges from [0..8]. Local orientation information is obtained by convolution with oriented Gabor pyramids $O(\sigma, \theta)$, where [0..8] is the scale and [0°, 45°, 90°, and 135°] is the preferred orientation.

A Winner-Take-All algorithm determines the location of the most salient feature in the calculated saliency map to be the target of the next attention shift. An Inhibition-Of-Return (IOR) mechanism is added to prevent immediate attention shifts back to the current feature of interest, to allow other parts of the object to be explored. In the case of an overt attention shift, the target is foveated after the commanded saccade, and a new retinal image is formed. The new image is fed into the module as input for the next learning iteration. A covert attention shift, on the other hand, will not foveate the

attended target, and therefore the subsequent retinal image input remains unchanged. Both overt and covert attention play an equal role in selecting local features, which are obtained when part of an object falls in the attention window before and after attention shifts. The resulting local features are fed into the input layer of the four-layer network.

2.2 Learning of Position-Invariant Local Features

The learning of position-invariant local features is based on the Clark-O'Regan approach ([3] [11]), which used a Temporal-Difference learning schema to learn the pair-wise association between pre- and post- motor visual input data, leading to a constant color percept across saccades. The learning takes place in the lower layers of the *learning module* from the input layer to the hidden layer. Our aim is to reduce the computational requirements of the Clark-O'Regan model while retaining the capability of learning position invariance. We make a modification to the learning rule, using temporal differences over longer time scales rather than just over pairs of successive time steps. In addition, we use a sparse coding approach to re-encode the simple neural responses, which reduces the size of the association weight matrix and therefore the computational complexity. The learning is triggered by the overt attention shift signal each time when a foveating saccade is committed.

We use a sparse coding approach [14] to reduce the statistical redundancies in the input layer responses. Let $\boldsymbol{I}$ denote the input layer neuronal responses. A set of basis functions *bf* and a set of corresponding sparsely distributed coefficients $\boldsymbol{S}$ are learned to represent $\boldsymbol{I}$:

$$I(j) = \sum_i S_i * bf_i(j) \Rightarrow I = bf * S \tag{1}$$

The basis function learning process is a solution to a regularization problem that finds the minimum of a functional $\boldsymbol{E}$. This functional measures the difference between the original neural responses $\boldsymbol{I}$ and the reconstructed responses $I' = bf * S$, subject to a constraint of sparse distribution on the coefficients $\boldsymbol{S}$:

$$E(bf, S) = \frac{1}{2} \sum_j [I_j - \sum_i S_i * bf_{ij}]^2 + \alpha \sum_i Sparse(S_i) \tag{2}$$

where $Sparse(a) = ln(1 + a^2)$.

The sparsely distributed coefficients $\boldsymbol{S}$ become the output of the encoded layer. A weight matrix $\boldsymbol{A}$ between the encoded layer and the hidden layer serves to associate the encoded simple neuron responses related to the same physical stimulus at different retinal positions. Immediately after an attention shift takes place, this weight matrix is updated according to a temporal difference reinforcement-learning rule, to strengthen the weight connections between the neuronal responses to the pre-saccadic feature and that to the post-saccadic feature.

The position-invariant neuronal response in the hidden layer $\boldsymbol{H}$ is represented by the following equation:

$$H = A * S \tag{3}$$

The updating is done with the following temporal reinforcement-learning rule:

$$\Delta A(t) = \eta * \{[(1-\kappa) * R(t) + \kappa * (\gamma * H(t) - \hat{H}(t-1))] * \hat{S}(t-1)\} \quad (4)$$

where

$$\Delta \hat{H}(t) = \alpha_1 * (H(t) - \hat{H}(t-1)) \quad (5)$$

$$\Delta \hat{S}(t) = \alpha_2 * (S(t) - \hat{S}(t-1)) \quad (6)$$

Here κ is a weighting parameter to balance the importance between the reinforcement reward and the temporal output difference between successive steps. The factor γ is adjusted to obtain desirable learning dynamics. The parameters η, α_1 and α_2 are learning rates with predefined constant values. The short-term memory traces, $\hat{H}$ and $\hat{S}$, of the neural responses in the hidden layer and the encoded layer, are maintained to emphasize the temporal influence of a response pattern at one time step on later time steps. The constraint of temporal perceptual stability requires that updating is necessary only when there exists a difference between current neural response and previous neural responses kept in the short-term memory trace. Therefore the learning rule incorporates a Hebbian term between the input trace and the output trace residuals (the difference between the current and the trace activity), as well as that between the input trace and the reinforcement signal.

Our approach is able to eliminate the limitations of Einhäuser *et al*'s model [4] without imposing an overly strong constraint on the temporal smoothness of the scene images. For example, in the case of recognizing a slowly moving object, temporal sampling results in the object appearing in different positions on the retina, which could cause temporal discontinuity in the Einhäuser *et al* model. But this will not affect the learning result of our approach because it employs a rapid overt attention shift to foveate the target object and only requires the temporal association between images before and after the attention shift.

2.3 Position-Invariant Representation of an Object Across Attention Shifts

Given that position-invariant representations of local features have been learned in the lower layers, an integration of local features of an object with fine details can be learned in the upper layers in the *learning module*. The representation of an object is learned in a temporal sequence as long as attention shifts stay within the range of the object. Here we assume that attention always stays on the same object during the recognition procedure of an object even in the presence of multiple objects. In our experiments this assumption was enforced by considering only scenes containing a single object.

A Winner-Take-All interaction ensures that only one neuron in the output layer wins the competition to actively respond to a certain input pattern. A fatigue process gradually decreases the fixation of interest on the same object after several attention shifts. This helps to explore new objects in the environment. For simplicity, in our experiments we restrict our scenes to contain only one object, we nonetheless implement the fatigue effect mechanism to simulate conditions such that an output layer neuron becomes fatigued after responding to the same object for a period of time. The fatigue effect is controlled by a Fixation-Of-Interest function $FOI(u)$. A value of $\boldsymbol{u}$ is kept for each output layer

neuron, in an activation counter initialized to zero. Each counter traces the recent neural activities of its corresponding output layer neuron. The counter automatically increases by 1 if the corresponding neuron is activated, and decreases by 1 until 0 if not. If a neuron is continuously active over a certain period, the possibility of its next activation (i.e., its fixation of interest on the same stimulus) is gradually reduced, and thus allows other neurons to be activated. A Gaussian function of u^2 is used for this purpose:

$$FOI(u) = \exp(-u^4/\sigma^2) \tag{7}$$

An output layer neural response C_0 is obtained by multiplying the hidden layer neural responses $\boldsymbol{H}$ with the integration weight matrix $\boldsymbol{W}$. C_0 is then adjusted by multiplying with $FOI(u)$, and is biased by the local estimation of the maximum output layer neural responses (weighted by a factor $\chi < 1$).

$$C' = C_0 * FOI(u) - \chi * \tilde{C}_0 \tag{8}$$

If C'_i exceeds a threshold, the corresponding output layer neuron is activated ($C_i = 1$).

The temporal integration of local features is accomplished by dynamically tuning the connection weight matrix between the hidden layer and the output layer. Neuronal responses to local features of the same object can be correlated by applying the temporal perceptual stability constraint that output layer neural responses remain constant over time. The constraint is achieved using a back-propagation term, where the short-term activity trace of the output neurons acts as a teaching credit to force the reduction of the temporal difference of neuronal responses between successive steps.

Given as input the position-invariant hidden layer neural responses $\boldsymbol{H}$ from the output of the lower layers, and as output the output layer neural responses $\boldsymbol{C}$, the weight matrix $\boldsymbol{W}$ is dynamically tuned using the learning rule as follows:

$$\Delta W(t) = \gamma * [\lambda * \hat{C}(t) - (1-\lambda) * (\hat{C}(t) - C(t))] * H(t) \tag{9}$$

where

$$\Delta\hat{C}(t) = \alpha * (C(t) - \hat{C}(t-1)) \tag{10}$$

The short-term activity trace $\hat{C}$ acts as an estimate of the output layer neuron's recent responses. The first term of the learning rule emphasizes the Hebbian connection between the output neuronal activity trace and the current hidden layer neuronal response; and the second term is the back-propagation term, which drives the updating of the weights towards constant output. The factor λ balances the importance between the Hebbian connection and the temporal continuity constraint.

3 Simulation and Results

In our model, position invariance is achieved when a set of neurons can discriminate one stimulus from others across all positions. Furthermore, the neural response should remain constant while attention shifts over the same object. We refer to "a set of neurons", as our representation is in the form of a population code, in which more than one neuron may exhibit a strong response to one set of stimuli. Between each set of neurons there might be some overlap, but the combination of actively responding neurons are unique, and can therefore be distinguished from each other.

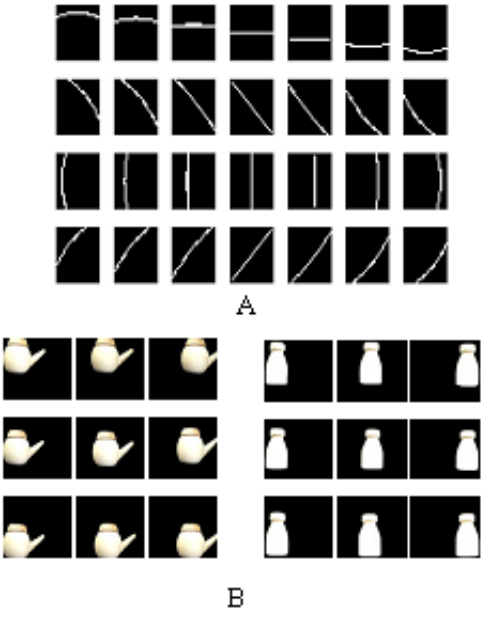

Fig. 2. Training data set of computer-simulated retinal images of lines (A) and image sequences of two real objects (B)

3.1 Demonstration of Position-Invariant Representation of Local Features

To demonstrate the process of position-invariant local feature extraction, we focus on the lower three layers: the input layer, the encoded layer and the hidden layer. The expected function of these layers is to produce constant neuronal responses to local features of an object at different positions after learning. We use two different test sets of local features as training data at this stage, a set of computer-generated simple features such as oriented lines at different position in a wiping sequence, and a set of computer-modified picture of real objects.

We first implemented a simplified model that has 648 input layer neurons, 25 encoded layer neurons and 25 hidden layer neurons for testing with the first training data set. The receptive fields of the input layer neurons are generated by Gabor functions over a 9x9 grid of spatial displacements each with 8 different orientations evenly distributed from 0 degree to 180 degree.

The first training image set is obtained by projecting straight lines of 4 different orientations ([0 °, 45 °, 90 °, and 135 °]) through a pinhole eye model onto 7 different positions of a spherical retinal surface. The simulated retinal images each have a size of 25x25 pixels. Figure 2A shows the training data set.

It was found in our experiment that some neurons in the hidden layer responded more actively to one of the stimuli regardless of its positions on the retina than to all other stimuli, as demonstrated in Figure 3. For example, neuron #8 exhibits a higher

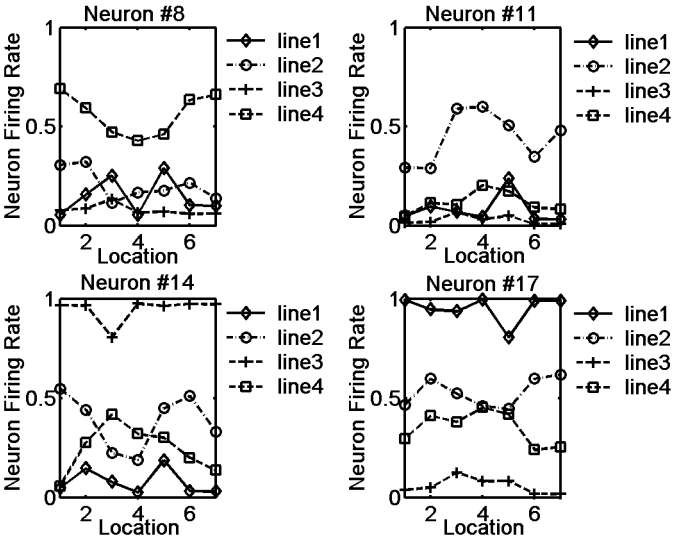

Fig. 3. Neural activities of the four most active hidden layer neurons responding to computer simulated data set at different positions

firing rate to line #4 than to any of the other lines, while neuron #17 responds to line #1 most actively. The other neurons remain inactive to the stimuli, which leave possible space to respond to other stimuli in the future.

It was next shown that the value of the weighting parameter κ in Equation 4 had a significant influence on this sub-module performance. To evaluate the performance, the standard deviation of activities of the hidden layer neurons are calculated when the sub-module is trained with different values of κ(=0, 0.2, 0.5, 0.7 and 1). The standard deviation of the neural activities is calculated over a set of input stimuli. The value stays low when the neuron tends to maintain a constant response to the temporal sequence of a feature appearing at different positions. Figure 4 shows the standard deviation of the firing rate of the 25 hidden layer neurons with different values of κ. The standard deviation becomes larger as κ increases. This result shows that the reinforcement reward plays an important role in the learning of position invariance. When κ is near 1, which means the learning depends fully on the temporal difference between stimuli before and after a saccade, the hidden layer neurons are more likely to have non-constant responses. Although there seems no much difference on the performance when the values of κ become very low, the reason why we keep the term of the temporal perceptual stability constraint in the learning rule is that this constraint forces the learning towards a constant state more quickly therefore makes the learning speedy. Choosing a proper value of κ is to balance the tradeoff between the performance and the speed. In practice we choose a non-zero but a rather small value of κ, for example $\kappa = 0.2$ for the most of the simulations we have run.

In our second simulation we tested image sequences of real world objects, such as a teapot and a bottle (Figure 2B). The images were taken by a digital camera with the

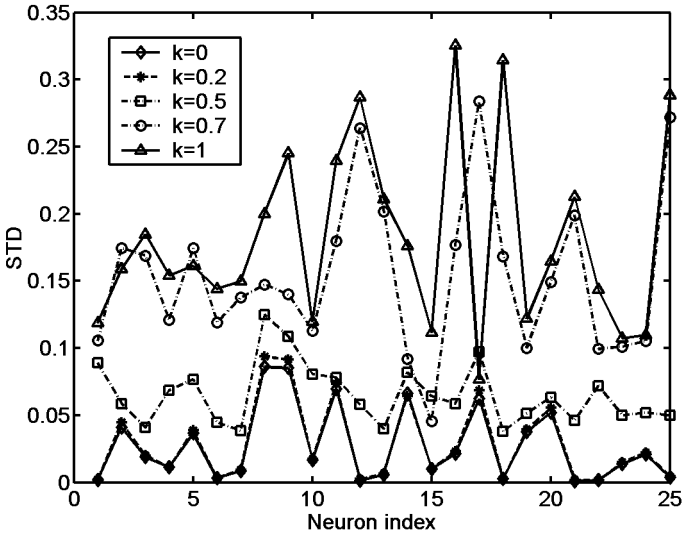

Fig. 4. Comparisons of position-invariant sub-module performance with varied weighting parameter κ ($\kappa = 0, 0.2, 0.5, 0.7, 1$), using a measurement of standard deviation of each neuronal response to a stimulus across different positions

center of its lens at nine different positions. Each image has a size of 64x48 pixels. The number of neurons in the encoded layer and the hidden layer has been increased from 25 to 64 from the numbers used in the previous experiment. This was required because the size of the basis function set to encode the sparse representations should also increase as the complexity of the input images increases.

Figure 5 shows the neural activities of the four most active neurons in the hidden layer when responding to the two image sequences of a teapot and a bottle respectively. Neurons #3 and #54 exhibit relatively strong responses to the teapot across all nine positions, while neuron #27 mainly responds to the bottle. We also have neuron #25 with strong overlapping neural activities to both stimuli. Satisfying our definition of position invariance, the sets of neurons that have relatively strong activities are different from each other.

3.2 Demonstration of Position-Invariant Representation of Objects

As the early learning process of integration is essentially random and has no effect on the later result, we use a gradually increasing parameter to adjust the learning rate of integration. This parameter can be thought of as an evaluation of the gained experience at the basic learning stage. The value of this parameter is set near 0 at the beginning of the learning, and near 1 after a certain amount of learning, at which point the position-invariant extraction process is deemed to have gained sufficient confidence in its experience on extracting position-invariant local features.

For simplicity, in this experiment we use binary images of basic geometrical shapes such as rectangles, triangles and ovals. These geometrical shapes are, as in the previous

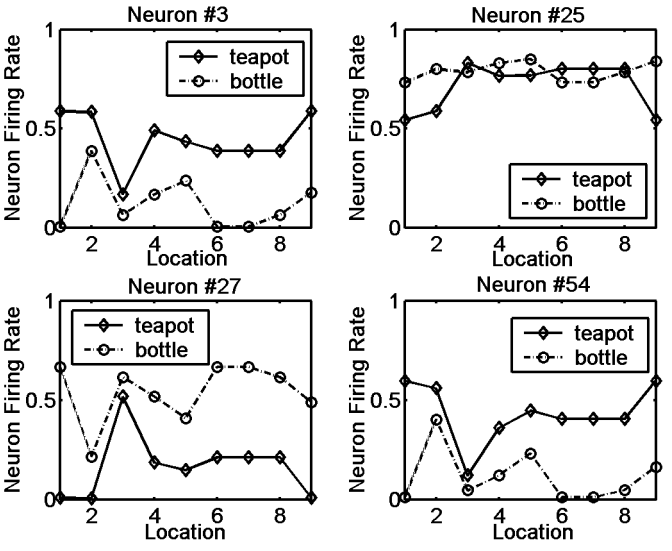

Fig. 5. Neural activities of the four most active hidden layer neurons responding to two real objects at different positions

experiment, projected onto the hemi-spherical retinal surface through a pinhole. Their positions relative to the fovea change as a result of saccadic movements.

Here we use an equal-weighted combination of intensity contrast and orientation contrast to compute the saliency map, as they are the most important and distinct attributes of the geometrical shapes we use in the training. A Winner-Take-All mechanism is

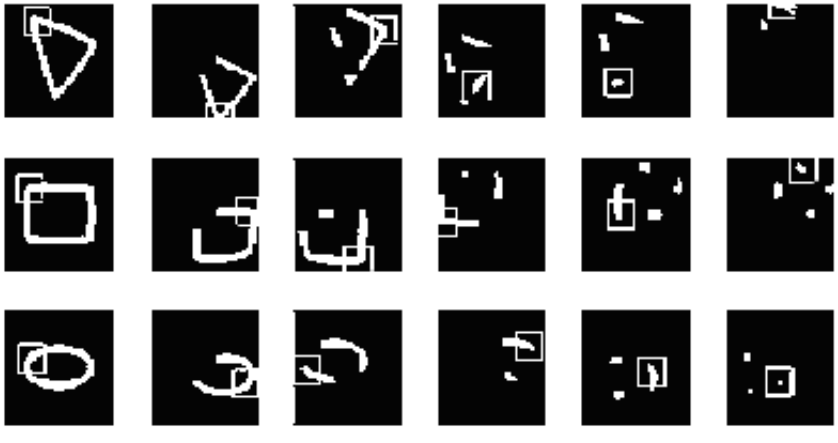

Fig. 6. Dynamically changing saliency maps for three geometrical shapes after the first 6 saccades following an overt attention shift. The small bright rectangle indicates an attention window centered at the most salient point in the saliency map

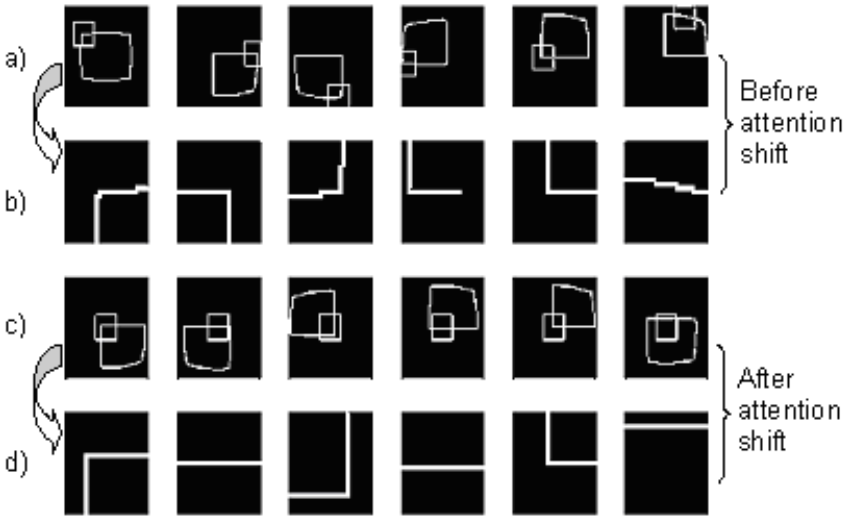

Fig. 7. Local features of a rectangular shape before (b) and after (d) an overt attention shift. a) and c) show retinal images of the same rectangle at different positions due to overt attention shifts

employed to select the most salient area as the next fixation target. After a saccade is performed to foveate the fixation target, saliency map is updated based on the newly formed retina, and a new training iteration begins. Figure 6 shows a sequence of saliency maps dynamically calculated from retinal images of geometrical shapes for a sequence of saccades. Local features of an object are obtained by falling in the attention window through these saliency maps as shown in Figure 7. Figure 7b and 7d show a sequence of pre- and post-saccadic local features of the retinal images of a rectangular shape falling in a 25x25 pixel attention window respectively.

Position-invariant representations of these local features are achieved from the previous learning steps in the lower layers, and they are fed into the integration procedure for the learning of invariance across attention shifts.

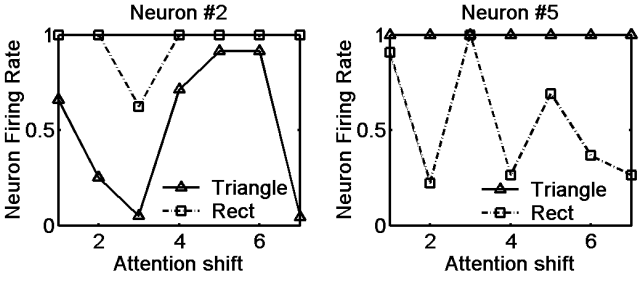

Fig. 8. Neural activities of the two most active output layer neurons responding to two geometric shapes across attention shifts

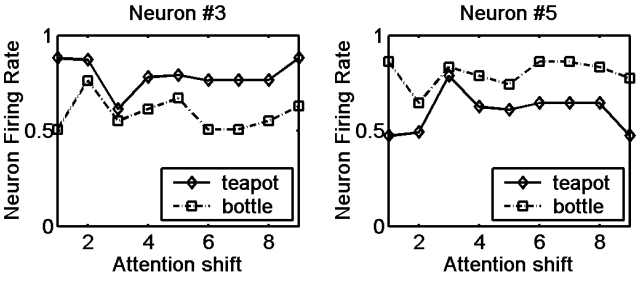

Fig. 9. Neural activities of the two most active output layer neurons responding to two objects (a teapot and a bottle) across attention shifts

We show in Figure 8 some of the output layer neural responses (neuron #2 and neuron #5) to two geometrical shapes: a rectangle and a triangle. Neuron #2 responds to the rectangle more actively than to the triangle, while neuron #5 has higher firing rate to the triangle than to the rectangle. Figure 9 shows another experiment result on two real objects, where output layer neuron #3 and neuron #5 have stronger response to a teapot and a bottle respectively.

4 Discussions

Our proposed approach is different from the dynamic routing circuit of Olshausen et al. [13], which forms position- and scale- invariant representations of objects for recognition via attention shifts. Firstly, attention shift plays different roles in both models. In Olshausen's model attention shift is merely for feature segmentation and selection. In our approach, attention shift also provides a reinforcement signal for learning, and there is a constraint on the learning rule that the output should remain constant when attention stays on the object. Secondly, Olshausen's model does not take into account image distortions due to the non-uniformity of the retina and the nonlinearity of projection onto the hemi-spherical retina when foveating eye movements take place. The object to be recognized undergoes simply translation and rescaling and therefore having little shape dissimilarity at each position and scale, while in practice distortions might lead to a greatly different appearance when the object is projected on the periphery. Our model is designed to consider the position-related distortions in the retinal input, leading to a position-invariant representation of an object. The overt attention shifts employed in our model brings an object in the periphery into the fovea, and correlates the before- and after- saccade retinal information to form a canonical representation regardless of its distortion.

The current approach is trained on one object per image. A further work has to be done on multiple objects in the scene with different background. To keep attention mostly staying on the same object during the learning, we can modify the formation of the saliency map by introducing a top-down stream which has preference to the neighborhood of the current focus of attention. It is true that there can be attention shifts that cross between objects. In practice, however, a given object will appear in conjunction

with a wide range of different objects from time to time. Thus, during learning in a complex uncontrolled environment, there will be few persistent associations to be made across inter-object attention shifts. For intra-object attentions shifts, on the other hand, the features will, in general, be persistent, and strong associations can be formed. In our view (which we have not tested) is that inter-object attention shifts will only slow down learning and not prevent it. The other methods for position-invariance also have problems with multiple objects, and they assume some form of scene segmentation to have been done.

The work described here mainly concerns learning of a constant neural representation of objects with respect to the position variance and the corresponding position-related distortions. The representation can be further applied to the task of object recognition using some supervised learning rule to associate the neural coding of an object with a certain object class. While recognizing an object across a sequence of attention shifts, the activities of the output neurons can be viewed as an evaluation of the confidence on a certain object. Although during the attention shifts, only parts of the object can be focused and some of the parts might coincident with other objects, we hypothesize that for most objects the set of their salient points are different. Therefore a temporal trace of the output neural activities over attention shifts is likely to represent an object as a whole uniquely in the form of its neural coding and can distinguish one from the others.

5 Conclusions

In this paper we have presented a neural network model that achieves position invariance incorporated with visual-related self-action signals such as attention shifts. Attention shifts play an important role in the learning of position invariance. Firstly, attention shifts are the primary reason for images of object features to be projected at various locations on the retina, which enables the learning focus on position invariance. Secondly, attention shifts actively select local features as input to the learning model. Thirdly, the motor signals of attention shift are used to gate the learning procedure.

We implemented a simplified version of our model and tested it with both computer-simulated data and computer-modified images of real objects. In these tests local features were obtained from retinal images falling in the attention window by an attention shift mechanism. The results show that our model works well in achieving both position invariance, regardless of retinal distortions.

References

1. Bandera, C., Vico, F., Bravo, J., Harmon, M., and Baird, L., Residual Q-learning applied to visual attention, *Proceedings of the 13th International Conference on Machine Learning*, (1996) 20–27
2. Becker, S., Implicit learning in 3D object recognition, The importance of temporal context. *Neural Computation* **11(2)**, (1999) 347–374
3. Clark, J.J. and O'Regan, J.K., A Temporal-difference learning model for perceptual stability in color vision, *Proceedings of 15th International Conference on Pattern Recognition* **2**, (2000) 503–506

4. Einhäuser, W., Kayser, C., König, P., and Körding, K. P., Learning the invariance properties of complex cells from their responses to natural stimuli, *European Journal of Neuroscience* **15**, (2002) 475–486
5. Földiák , P., Learning invariance from transformation sequences, *Neural Computation* **3** (1991) 194–200
6. Henderson, J.M., Williams, C.C., Castelhano, M.S., and Falk, R.J., Eye movements and picture processing during recognition, *Perception and Psychophysics* **65(5)**, (2003) 725–734
7. Itti, L., Koch, C. and Niebur, E., A model of saliency-based visual attention for rapid scene analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20(11)**, (1998) 1254–1259
8. Kikuchi, M., and Fukushima, K., Invariant pattern recognition with eye movement: A neural network model, *Neurocomputing* **38-40**, (2001) 1359–1365
9. Koch, C., and Ullman, S., Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology* **4**, (1985) 219–227
10. Körding, K. P. and König, P., Neurons with two sites of synaptic integration learn invariant representations, *Neural Computation* **13**, (2001) 2823–2849
11. Li, M. and Clark, J.J., Sensorimotor learning and the development of position invariance, poster presentation at *the 2002 Neural Information and Coding Workshop*, Les Houches, France, (2002)
12. Minut, S., and Mahadevan, S., A reinforcement learning model of selective visual attention, *AGENTS2001*, (2001) 457–464
13. Olshausen, B.A., Anderson C.H., Van Essen D.C., A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information, *The Journal of Neuroscience* **13(11)**, (1993) 4700–4719
14. Olshausen, B. A. and Field, D. J., Sparse coding with an overcomplete basis set: A strategy employed by V1?, *Vision Research* **37**, (1997) 3311–3325