Biomorphic networks: Approach to invariant feature extraction and segmentation for ATR*

A. S. Baek and N. H. Farhat
Department of Electrical Engineering
University of Pennsylvania
Philadelphia, PA 19104

ABSTRACT

Invariant features in two dimensional binary images are extracted in a single layer network of locally coupled spiking (pulsating) model neurons with prescribed synapto-dendritic response. The feature vector for an image is represented as invariant structure in the aggregate histogram of interspike intervals obtained by computing time intervals between successive spikes produced from each neuron over a given period of time and combining such intervals from all neurons in the network into a histogram. Simulation results show that the feature vectors are more pattern-specific and invariant under translation, rotation, and change in scale or intensity than achieved in earlier work. We also describe an application of such networks to segmentation of line (edge-enhanced or sihlouette) images. The biomorphic spiking network's capabilities in segmentation and invariant feature extraction may prove to be, when they are combined, valuable in Automated Target Recognition (ATR) and other automated object recognition systems.

Keywords: Spiking neural networks, biomorphic neuron, feature extraction, distortion invariance, segmentation, automated object recognition

1. INTRODUCTION

Invariant feature extraction plays a central role in designing automatic object recognition systems, whose applications permeate a diverse range of fields that include radar, sonar, optical character recognition, autonomous vision machines for use in space exploration, robotics, and manufacturing. The human visual system recognizes objects remarkably well from two dimensional images cast on the retina, even when these images are distorted by various factors such as change in the distance and orientation of the object, light illumination level, shading and background clutter. Understanding and duplicating this exceptional invariant recognition ability of the visual system is valuable for successful development of advanced automatic target recognition (ATR) systems.

To this end artificial neural network models based on neuronal firing rate have been explored and applied in the past, with various degrees of success, to both the feature extraction and the recognition aspects of ATR [1]. It is however becoming increasingly clear that the temporal aspects of brain function manifested in the spiking nature of neural activity and the relative timing of spikes may play an important role in distortion invariant feature-extraction, feature-binding and other higher-level brain functions.

In this paper we will present methods for invariant feature extraction and image segmentationin using a network of biomorphic spiking neurons that extends and improves on earlier published work. Invariant features in two dimensional binary images are extracted in a single layer network of locally coupled spiking (pulsating) neurons that include synapto-dendritic processing. The feature vector for an image emerges as invariant structure in the aggregate histogram of interspike intervals, which is formed by computing time intervals between successive spikes produced from each neuron over a given period of time and combining such intervals from all neurons in the network into a histogram. The simulation results presented here show that the feature vectors are object-specific and highly invariant under translation, rotation, and change in scale or intensity. The results suggest that the combined process

* Based in part on an oral presentation given given at the 1997 OSA Annual Meeting, Long Beach, CA Oct. 1997.

of the segmentation and feature extraction from segmented images may provide useful basis in designing ATR and other autonomous recognition systems.

Current techniques for distortion invariant pattern recognition use integral transforms, algebraic moments, or perceptron-like neural networks with learning algorithms [2]. These techniques demand, however, extensive computational resource and are susceptible to spatial noise in the input image. Recently, a novel pulse-coupled neural network, which could generate invariant signatures of images of simple canonical patterns under translation, rotation, scale and intensity changes, was described in [3]. Although our method is an extension of this earlier work in that it also uses a pulse-coupled neural network and employs concepts developed in [3] and [4], there is a fundamental difference between the network dynamics in the two methods. The method presented in [3] is based on the cortical model of synchronization of neural temporal activity developed in [4] but uses moderate-to-weak linking strengths to induce phase-locked firing states of neurons. Also in [3], the spatial structure of the test patterns (a cross and a tee), made of blocks with each block containing 11×11 pixels and a distinctive gray-scale intensity level, was encoded in the phase-locking pattern of output spikes from neurons. In contrast, the method described here is less "corticomorphic". It is based on firing rate encoding, which is predominantly observed in sensory nervous systems, rather than phase-locking, and is suited for processing binary line images. This different encoding scheme arises, as explained below, from the opposite relationship between the time constants of the synaptic response function and the membrane potential (pulse generator) that we use as compared to [3]. As a result, the method appears to produce invariant feature vectors that are more input pattern specific than in [3], which is a desirable prerequisite for accurate classification. Also, its ability to extract invariant features of binary line images is an attractive attribute because edge-enhancement and line extraction are standard operations in pattern recognition.

In Section 4, we demonstrate the ablity of the spiking neural networks in segmentation of line images (eg. silhouettes or edge enhanced images) of model objects. Once characteristic segments of a line image are obtained, each segment may be processed by a biomorphic spiking network, similar to the one described in Section 2, to produce a set of histograms of interspike intervals, which contain invariant features that are specific to the line image. The resulting set of invariant features, which belong to the respective features of the image, can then be regarded as a composite invariant feature vector that represents the entire line image.

2. THE NETWORK

In the invariant feature extraction system presented here, a basic image that consists of a 12×12 binary pixel array is fed to a single layer network that consists of spiking neurons arranged in the 12×12 array format. The model neuron used in the network is of the integrate-and-fire (I&F) variety , similar to that used in [3]-[5], and is regarded as biomorphic in that it contains a simplified model of the dendritic-tree processing in biological neurons, as explained below.

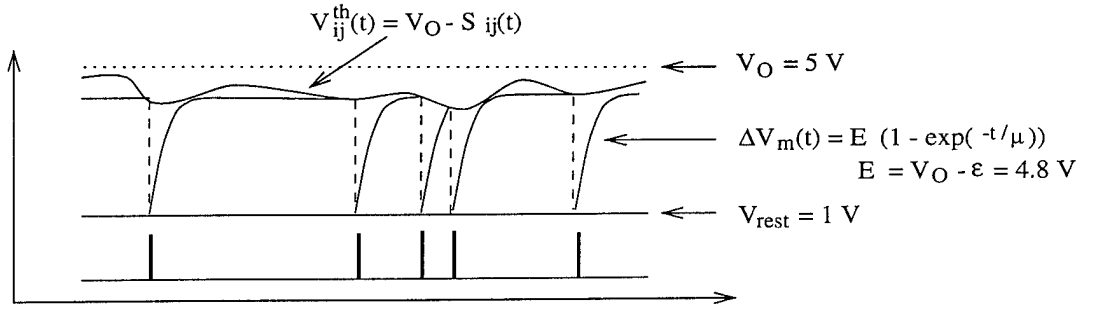
The dynamics of the neuron can be summarized as follows, refering to Fig.1. When the exponentially rising membrane potential, $\Delta V_m(t)$, of the neuron at (i,j) reaches the time-varying threshold, $V_{ij}^{th}(t)$, it instantaneously drops to the resting membrane potential V_{rest} and at the same time a spike (action potential) is generated [6]. The time-varying threshold is represented as

$$V_{ij}^{th}(t) = V_0 - U_{ij}(t) \tag{1}$$

where V_0 is constant and $U_{ij}(t)$, as defined below in Eqn. (3), is a signal that results from the joint actions of synaptic inputs and the intensity of the image. This process of exponential rise and instantaneous drop of the membrane potential repeats, generating successive spikes whose timing is modulated by the threshold voltage. Note that lowering the threshold causes the neuron to fire faster, because it takes less time to build the membrane potential from the resting potential to the threshold.

In an effort to emulate the signal processing that occurs in the dendritic tree of biological neurons, the model neuron is given a simplified homogeneous synapto-dendritic response function modeled as an impulse response function, which approximates the change in membrane potential (depolarization) at the neuron's hillock (i.e, the site of output spike generation), in response to a single spike on its dendritic tree. We use an exponentially decaying function for the synapto-dendritic response, given by

$$h(t) = h_o \exp(-t/\tau) \tag{2}$$



$S_{ij}(t) = I_{ij}(1 + \beta P_{ij}(t))$

$P_{ij}(t) = ST_{ij} * h(t)$

$h(t) = h_o \exp(-t/\tau)$

$\mu = 1$ msec (membrane time constant)

$\tau = 12$ msec (synapto-dendritic time constant)

Figure 1. The dynamics of the biomorphic spiking neuron and parameters used in numerical simulations.

where τ is the time constant of the response. Also, the model neuron in the network receives the intensity value of the image at the corresponding coordinates, which modulates the signal produced from synaptic inputs in a nonlinear manner similar to that in [3] and [4], as further explained below.

Using the same notation as in [3] and [4], the signal, $U_{ij}(t)$, which is produced from a nonlinear combination of the external input and synaptic inputs, is determined by Eqn. (3) with i and j representing the coordinates of both the pixel and the neuron in the 12×12 array. The signal $P_{ij}(t)$, as defined in Eqn. (4), is a convolution of the synapto-dendritic response function and the input spike train impinging on the neuron at (i, j). The input spike train is represented below as a sum of delta functions with m and n representing the coordinates of the sending neuron and k representing the spike generation time.

$$S_{ij}(t) = I_{ij} \times (1 + \beta P_{ij}(t)) \tag{3}$$

$$P_{ij}(t) = ST_{ij} * h(t) \tag{4}$$

$$ST_{ij}(t) = \sum_{mnk} \delta(t - t_{mn}^k) \tag{5}$$

The signal $S_{ij}(t)$ modulates the threshold voltage of the neuron at (i, j), as shown in Eqn. (1), and this will determine the firing times. The departure of the membrane potential V_m from the resting value V_{rest} can be expressed as:

$$\Delta V_m(t) = E(1 - \exp(-\frac{t}{\mu})) \tag{6}$$

$$E = V_o - \epsilon \tag{7}$$

where V_o and ϵ are given values of 5 and 0.2 volts, respectively in the simulations below. When a spike is elicited, the membrane potential V_m instantaneously drops to the resting potential V_{rest} and begins to rise exponentially towards the time-varying threshold, $V_{ij}^{th}(t)$ as depicted in Fig.1.

Attention is drawn to the relationship between the time constant μ of the exponential membrane build-up and the time constant τ of the synapto-dendritic response function. In contrast to the practice in [3] in which the time constant of the synapto-dendritic response is shorter than that of the membrane dynamics (referred to as pulse generator in [3]), the time constant τ of the synapto-dendritic response here is longer than the time constant μ of the membrane potential build-up. This alteration of the relation between the time constants produces a markedly different network behavior in that the prolonged synapto-dendritic response causes slow depressions in the threshold of a neuron, affecting its firing rate. In our simulations, the time constant τ of the synapto-dendritic response function is $12msec$, while that of the membrane potential build-up is $1msec$. In contrast the time constants in [3] are 1 for synaptic response and 5 for the membrane threshold (pulse generator) (units were not specified).

Another important feature of the model neuron used here, which is adopted from [3] and [4], is the nonlinear combination of the extrinsic input I_{ij} and the intrinsic input from other neurons $P_{ij}(t)$, as defined by Eqn. (3). The

pixel intensity of the image modulates the signal produced from synaptic inputs. The strength of the modulation is controlled by the constant β. Especially, when I_{ij} is zero, the neuron at (i, j) can not fire, because $U_{ij}(t)$ then becomes zero, according to Eqn. (3), and the neuron operates in the sub-threshold mode (the membrane potential saturates before reaching the threshold voltage). The value of $I_{ij} = 0$ effectively nullifies any effects of synaptic inputs on the threshold, forcing the neuron to be "silent". This nonlinear operation is similar to the biological situation in which inhibitory inputs proximal to the soma (cell body) may effectively nullify excitation by inputs on distal synapses [7]. In a complex morphological dendritic tree, this type of a selective AND-NOT like operation can effectively decouple subunits of the tree.

The network architecture used is shown in Fig. 2(a). The neurons are arranged in a 12×12 array and a sample neuron in the lower right side of the array is depicted showing its synaptic connections from its 8 nearest neighboring neurons. Interspike intervals of the spike train from each neuron are computed in the local interspike interval analyzers and the results are combined to form the aggregate interspike interval histogram. Shown in Fig. 2(b) is the schematic of an isolated neuron with the two types of inputs, the instrinsic synaptic inputs from the neighboring neurons and the extrinsic pixel intensity value from the image.

3. SIMULATION RESULTS

The simulations results are shown in Figs. 3, 4, and 5 with the input images and the associated aggregate (interspike interval histograms) ISIHs shown in the left and the right columns, respectively. Figure 3(a) shows the simulation result with the original image of the tee. The input image in Fig. 3(b) is a scaled, 90-degree rotated, and translated version of the original tee image. Figures 3(c) and (d) show the input images of the cross and the associated output ISIHs. Already from these simulations, one can clearly see the differences between the histograms produced from the tee images and from the cross images and the invariance of the histograms with image distortions. In Figs. 4(a) and (b) for a rectangle, the histograms are drastically different from the previous cases of the tee and the cross, highlighting the pattern-specificity, and the invariance is well retained for the rectangle in the distortions of scale, rotation, and translation. As reader might have noticed, the angles used in the simulations for the rotational invariance are confined to the multiples of 90 degrees, due to the limited spatial resolution. At higher spatial resolution, this restriction on the rotation angle can be relaxed and the rotation invariance for in-between angles will be achieved. The invariance with scale for a triangle is shown in Fig. 4(c) and (d). We have also examined invariance with intensity for the triangle, as demonstrated in Figs. 5 (a) and (b). The uniform intensity of the triangle used for Figs. 5 (a) is 10 percent higher than that used for Figs. 5 (b). The characteristic form of the histogram for the triangle is seen to be retained but experiences a shift towards lower time interval region. The shift comes from the fact that the higher uniform image intensity lowers the thresholds of the neurons, raising the firing rate. On a relevant note, we have an indication from later simulations that using a synapto-dendritic response function that is more biomorphic, i.e., approximates the biological response more accurately [8], results in higher invariance to the distortions, especially the intensity change. Results from these simulations are not reported here due to space limitation.

The invariant structure in ISIH is closely related to characteristic local features of the input pattern. For instance, the main feature of the triangle is the three corners. The neurons residing in these corners of the triangle and the vicinities have higher firing rates that are responsible for invariant structure in the ISIH, since they are connected to more neighboring neurons with higher activity than the neurons in the middle of edges. Such analysis carried out by the network is meaningful for syntactical information processing in that the input image can be implicitly segmented into sections that contain characteristic structural information. It is, however, important to remember that due to the convoluted nature of inter-neuron interactions through pulse-coupling, the invariant structure in the ISIH is also affected indirectly by activity of neurons that are beyond the nearest neighboring connections. Although this distal interaction is much weaker than the local interaction, it still seems to take a part in determining the shape of the ISIH.

4. SPIKING NETWORK FOR SEGMENTATION

A preliminary study of the ability of biomorphic spiking networks to segment line images (i.e., silhouettes or edge enhanced images) of model objects was carried out. The representative results given next suggest that combining spiking networks for segmentation with spiking networks for invariant feature extraction may offer a viable approach for generating invariant features for extended objects.

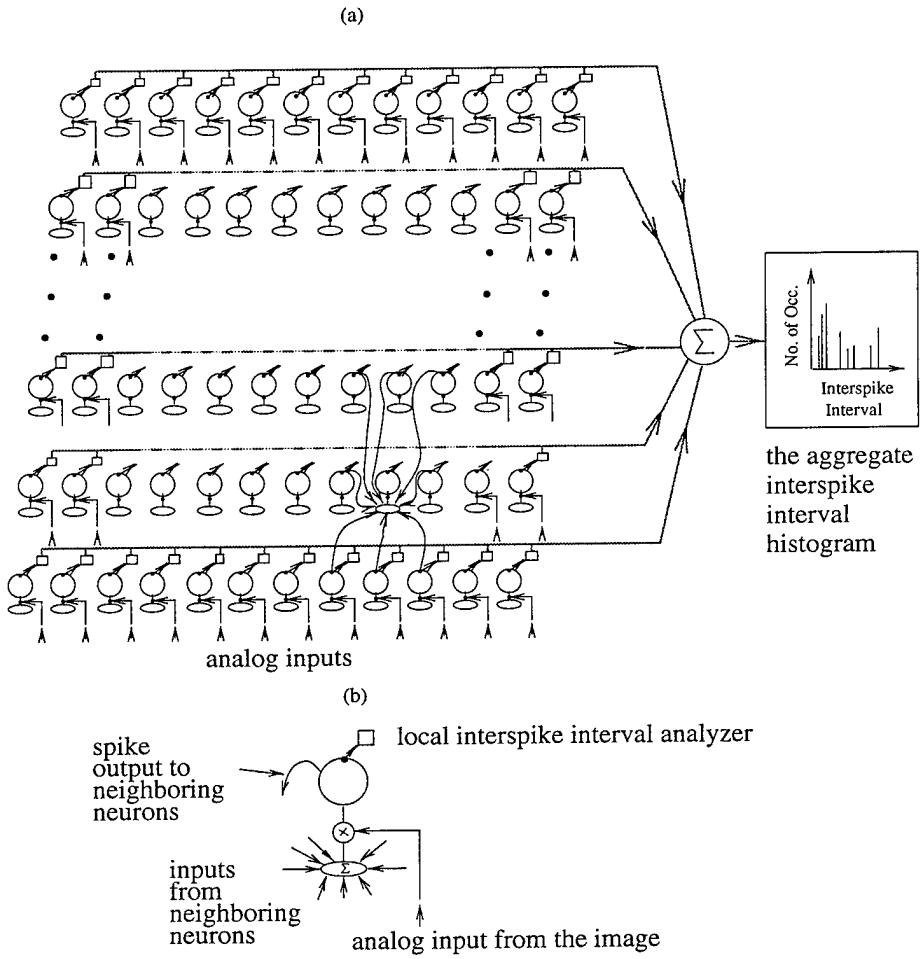


Figure 2. (a) Schematic of the network arranged in the $N \times N$ array. A set of exemplar synaptic connections of a neuron from its eight neighboring neurons is shown in the lower right side of the array, with the cursive lines representing the connections. Interspike intervals from each neuron are computed in the local interspike interval analyzer and sent to the aggregate interspike interval histogram. (b) Schematic of the neuron with two types of inputs: synaptic inputs from its neighboring neurons and an analog input of the image intensity at the neuron's location. The two types of the inputs are combined in a multiplicative manner to produce a signal that influences the timing of action potential generation and therefore the spiking output to the synaptically connected neurons.

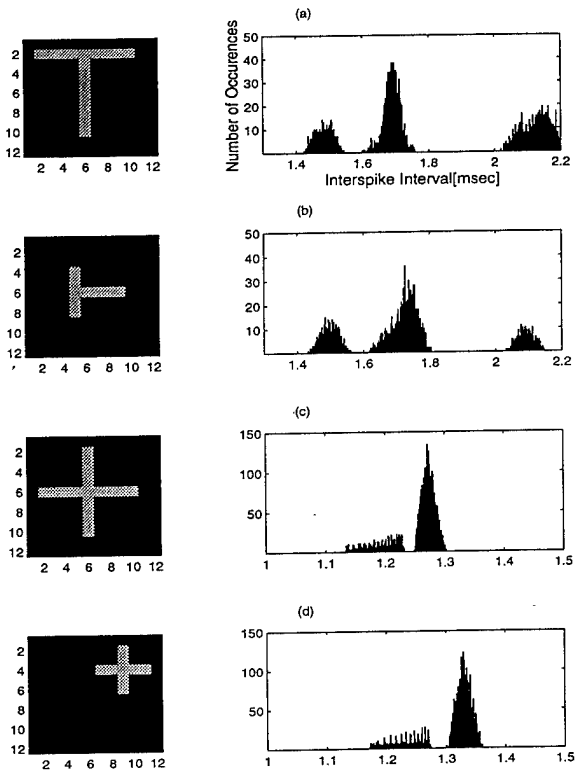


Figure 3. Simulation results with the tee and the cross demonstrating invariance with translation, rotation, and scale. The input images and the aggregate (interspike interval histograms) ISIHs from simulations are shown in the left and the right column, respectively. (a) The tee image, (b) A 90-degree rotated, scaled, and translated version of the tee in (a), (c) The cross image, (d) A scaled and translated version of the cross in (c). The time resolution of the ISIHs is 0.001 msec.

In order to obtain the input line images, test objects (two types of model aircrafts and one tank) were photographed using a CCD camera, then digitized and saved in gray-scale in 400×400 pixel format. Examples of such gray scale images are shown in Figs. 6(a), 7(a), and 8(a). The gray-scale images were then binarized and edge-detected to produce the line images as shown in Figs. 6(b), 7(b), and 8(b). In producing the line image, if the gray-scale image is smaller than 400×400, we padded the remaining pixels with zero intensity to maintain a consistent image size. Then, the line image was fed into the biomoprhic neural network, which has the same dimensions as those of the input image format. The neurons of the network used here employed an alpha function [8] for the synapto-dendritic response in place of the exponential function given by Eqn.3. Note that the neurons which receive zero intensity from the image do not fire, because the signal $S_{ij}(t)$ in Eqn. (3) is zero for all such neurons.

Figures 6(c), 7(c), and 8(c) present the simulation results with the neurons' firing rates given in gray scale. the synaptic weight matrix used in the simulation was circularly symmetric with a radius of 21 pixels, and had a uniform excitatory synaptic weight of one within the circle. In these plots, the brighter pixels indicate higher firing rates for the neurons located in those spots. We then varied a threshold firing rate and plotted only those pixels whose firing rate is above the threshold, as shown in Figs. 6(d), 7(d), and 8(d) to segment the line images into characteristic segments each of which can potentially be represented by an invariant ISIH with the aid of a network similar to that described in Section 2. The pronounced features in Fig. 5 6(d) are the engines in the rear, the tips of the wings, the angels at the functions of the wings and the main body, while in Fig. 7(d) they are the tail, the engines and the gun barrels on the wings, and the front tip. In Fig. 6(d), the biomoprhic network extracted (segmented) the most characteristic feature of tanks, the nozzle.

Figures 9 and 10 show the invariance of segmenting the line images of Aircrafts 1 and 2 under translation, rotation, and change of scale. The simulation results using images of Aircrafts 1 and 2 are shown with the input edge-enhanced images given on the left and the simulation results showing neurons with firing rates above a certain

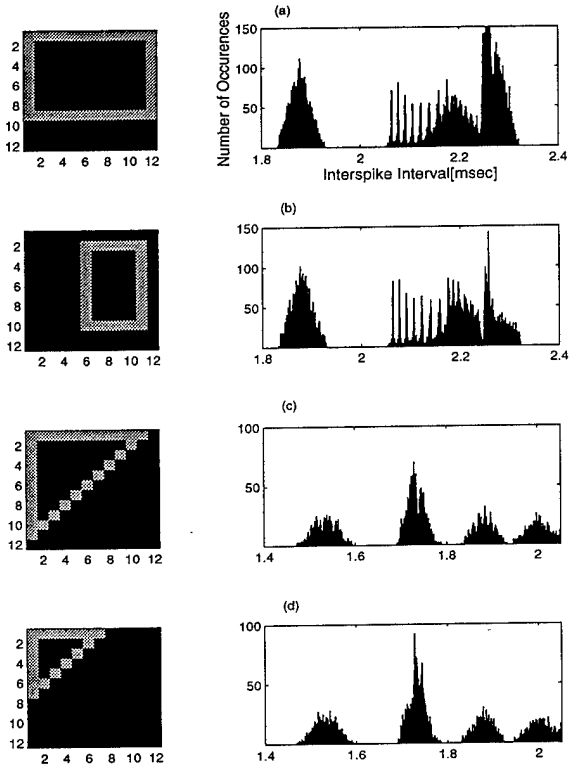


Figure 4. Simulation results with the rectangle and the triangle demonstrating invariance with invariance of translation, rotation, and scaling. The input images and the aggregate ISIHs from simulations are shown in the left and the right columns, respectively. (a) The rectangle image, (b) A 90-degree rotated, scaled, and translated version of the rectangle in (a), (c) The triangle image, (d) A scaled version of the triangle in (c).

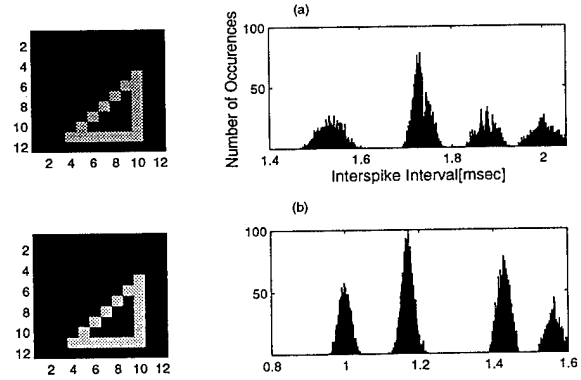


Figure 5. Simulation results with the triangle for intensity invariance. The input images and the aggregate ISIHs from simulations are shown in the left and the right columns, respectively. (a) The triangle image, (b) The triangle image with uniform intensity increased by 10 percent.

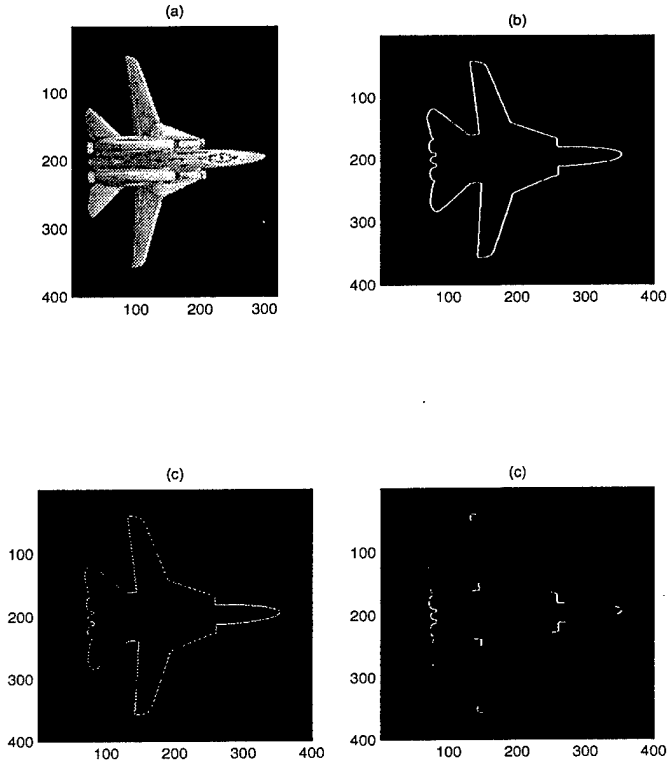


Figure 6. Segmentation of gray scale image of model Aircraft 1. (a) Gray scale image of model Aircraft 1, (b) The line image of the gray scale image in (a), (c) Firing rates of the biomorphic neurons with the brightest pixels representing the highest firing rates, (d) Plot of firing rates above a prescribed threshold. The plot shows characteristic segments of the aircraft, each of which maybe further analyzed by a feature extracting network to represent it by an invariant ISIH. The invariant histograms of the segments form an "invariant set", which can be used to represent the object and serve as an input to an ATR system.

threshold in the right. In these simulations, the circularly symmetric synaptic weight matrix has two concentric circles with radii of 48 and 51 pixels. The matrix elements inside the inner circle with radius of 48 pixels have a uniform excitatory synaptic weight of one, while those between the inner and the outer circles have a uniform inhibitory synaptic weight of negative one. We observed that the circularly symmetric weight with a large radius produces better rotation invariance. When the radius decreases, the synaptic matrix delineated over the rectangular array loses circular symmetry and becomes a square, due to the spatial resolution limitation. Also in the simulations shown in Figs. 9 and 10, the inhibitory weight is used to balance off the excitatory synaptic connections and avoid saturation of neurons' firings. The simulation results show that robust invariance under translation, rotation, and scale, is achieved.

In the above simulations involving model objects, as well as in the previous examples involving canonical objects, the geometrical complexities are encoded in firing rates of the neurons that reside in the corresponding locations. As we have seen, sections of the images that contain complicated structure become pronounced in the firing rate plot.

5. CONCLUSIONS

We have described a spiking neural network that can extract invariant feature vectors of two dimensional binary line images based on firing rate encoding rather than encoding by phase-locking described in [3]. The difference seems to lead to feature vectors or signatures for binary line images that are more pattern-specific and invariant under translation, rotation, and change in scale or intensity than achieved in [3]. Also the simpler neuronal and network structures used here can be advantageous in hardware implementation for real-time processing.

A preliminary study of the ability of biomophic spiking networks to segment line images (i.e., silhouettes or edge enhanced images) of model objects was also carried out. The results suggest that the the combined processing of

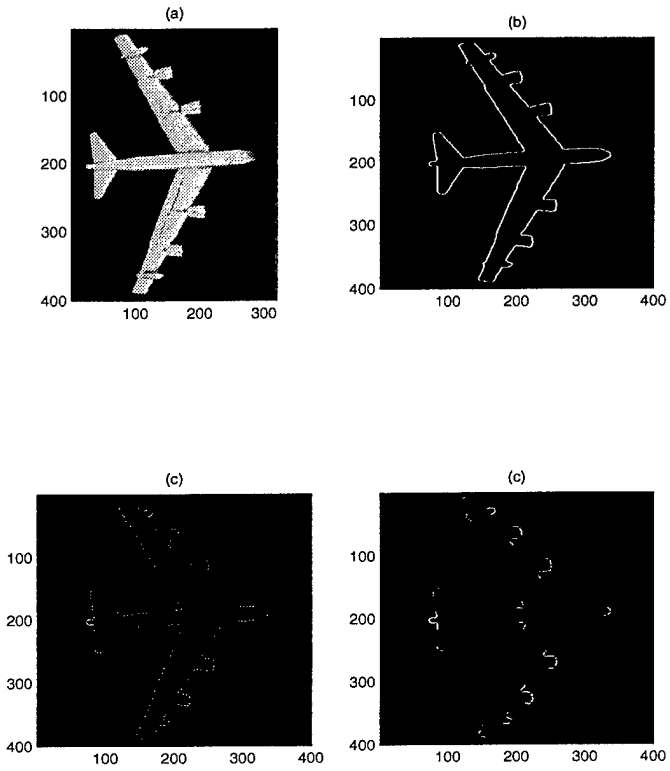


Figure 7. Segmentation of gray scale image of model Aircraft 2. (a) Gray scale image of model Aircraft 2, (b) The line image of the gray scale image in (a), (c) Firing rates of the biomorphic neurons with the brightest pixels representing the highest firing rates, (d) Plot of firing rates above threshold. The plot shows characteristic segments of the aircraft.

segmentation and feature extraction, each peformed by respective spiking networks, may provide a viable approach for generating invariant feature vectors for extended objects, which is worthy of further investigation.

ACKNOWLEDGMENTS

This research was supported by the Office of Naval Research under grant no. N0001-94-1-093P01 and in part by NASA Graduate Student Researchers Programs for one of us (A. Baek).

REFERENCES

1. H. Babri, "Neurodynamic networks for recognition of radar targets", Ph.D. dissertation, University of Pennsylvania, 1992.
2. J. Wood, "Invariant pattern recognition," *Pattern Recognition* **29**, pp.1, 1996.
3. J. L. Johnson, "Pulse-coupled neural nets: translation, rotation, scale, distortion, and intensity isgnal invariance for images," *Appl. Opt.* **108**, pp.6239, 1994.
4. R. Eckhorn, H. J. Reitboeck, M. Arndt, and P. Dicke, "Feature linking via synchronization among distributed assemblies: simulations of results from cat visual cortex," *Neural Comput.* **2**, pp.293 (1990).
5. N. H. Farhat, S. Y. Lin, and M. Eldefrawy, "Complexity and chaotic dynamics in a spiking neuron embodiment," *Proc. SPIE* **CR55**, pp.77, 1994.
6. J. G. Nicholls, A. R. Martin, and B. G. Wallace, *From Neuron to Brain*, Sinauer, Sunderland, MA, 1992.
7. C. Koch and T. Poggio, Trends in Neuroscience **3**, pp.455, 1983.
8. P. S. Anton, R. Granger, and G. Lynch, in *Single Neuron Computation* T. McKenna, J. Davis, and S. F. Zornetzer, eds. pp. 291, Academic, San Diego, CA, 1992d.

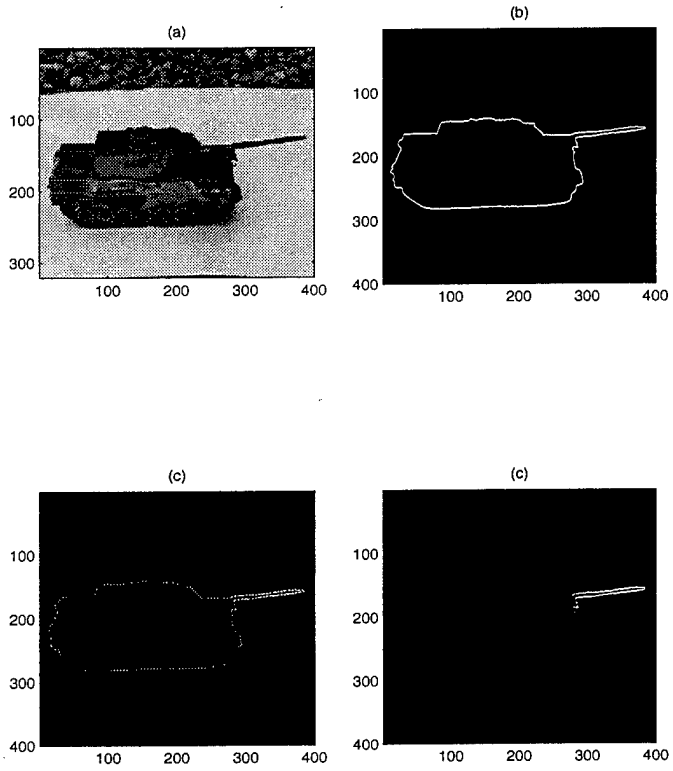


Figure 8. Segmentation of gray scale image of a model tank. (a) Gray scale image of a model tank, (b) The line image of the gray scale image in (a), (c) Firing rates of the biomorphic neurons with the brightest pixels representing the highest firing rates, (d) Plot of firing rates above a threshold. The plot shows a characteristic segment of the tank, the nozzle.

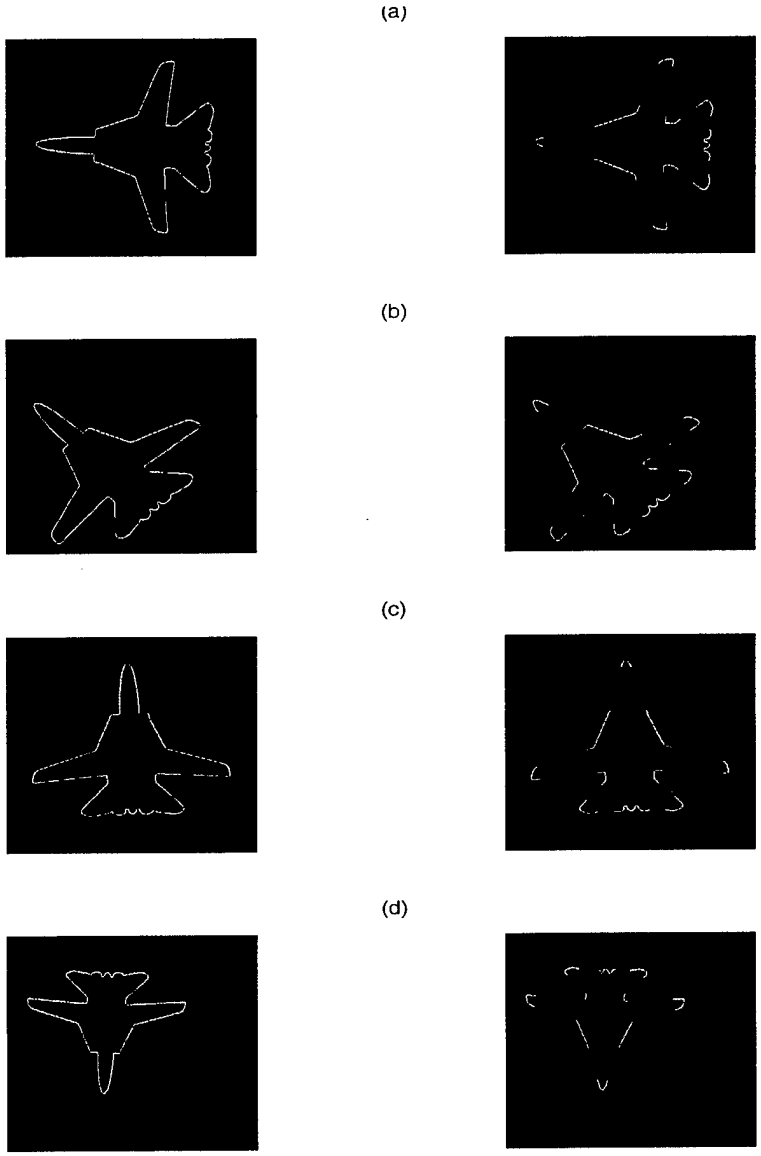


Figure 9. Invariant segmentation of edge-enhanced images of model Aircraft 1. The input images and the invariant segmented image formed by thresholding firing rates of the neurons are shown in the left and the right columns, respectively. (a) Segmentation of the original edge enhanced image of model Aircraft 1, (b) Segmentation of a 45-degree rotated and translated image of the original image shown in (a), (c) Segmentation of a 90-degree rotated image of the original, (d) segmentation of a 90-degree rotated, 80-percent scaled, and translated image of the original.

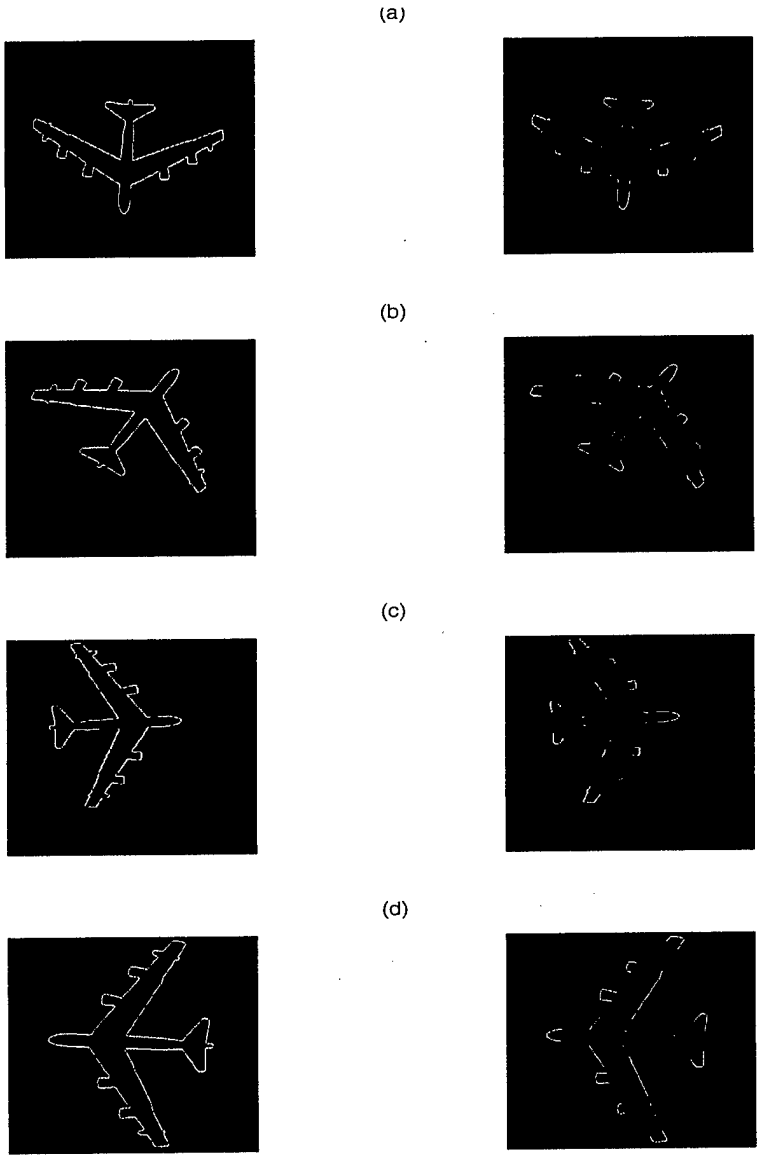


Figure 10. Invariant segmentation of edge-enhanced images of model Aircraft 2. The input images and the invariant features derived from thresholding firing rates of the neurons are shown in the left and the right columns, respectively. (a) Segmentation of the original edge enhanced image of model Aircraft 2, (b) Segmentation of a 135-degree rotated and translated image of the original image shown in (a), (c) Segmentation of a translated image of the original, (d) Segmentation of a 90-degree rotated and 120-percent scaled image of the original. Notice a change of scale of a segment will not affect the ability of a spiking network of the type described in Section 2 to extract an invariant IHIS.