Image Segmentation Based on Oscillatory Correlation

DeLiang Wang
Department of Computer and Information Science and Center for Cognitive Science,
The Ohio State University, Columbus, Ohio 43210, USA

David Terman
Department of Mathematics, The Ohio State University, Columbus, Ohio 43210, USA

We study image segmentation on the basis of locally excitatory, globally inhibitory oscillator networks (LEGION), whereby the phases of oscillators encode the binding of pixels. We introduce a lateral potential for each oscillator so that only oscillators with strong connections from their neighborhood can develop high potentials. Based on the concept of the lateral potential, a solution to remove noisy regions in an image is proposed for LEGION, so that it suppresses the oscillators corresponding to noisy regions but without affecting those corresponding to major regions. We show that the resulting oscillator network separates an image into several major regions, plus a background consisting of all noisy regions, and we illustrate network properties by computer simulation. The network exhibits a natural capacity in segmenting images. The oscillatory dynamics leads to a computer algorithm, which is applied successfully to segmenting real gray-level images. A number of issues regarding biological plausibility and perceptual organization are discussed. We argue that LEGION provides a novel and effective framework for image segmentation and figure-ground segregation.

1 Introduction

The segmentation of a visual scene (image) into a set of coherent patterns (objects) is a fundamental aspect of perception that underlies a variety of tasks, such as image processing, figure-ground segregation, and automatic target recognition. Scene segmentation plays a critical role in the understanding of natural scenes. Although humans perform it with apparent ease, the general problem of image segmentation remains unsolved in sensory information processing. As the technology of single-object recognition has become more and more advanced, the demand for a solution to image segmentation has increased since both natural scenes and manufacturing applications of computer vision are rarely composed of a single object.

Objects appear in a natural scene as the grouping of similar sensory features and the segregation of dissimilar ones. Sensory features are generally
taken to be local, and in the simplest case they may correspond to single pixels. To approach the problem of scene segmentation, three basic issues must be addressed: What are the cues that determine grouping and segregation? What is the proper representation for the result of segmentation? How are the cues used to give rise to segmentation?

Much is known about sensory cues that are important for segmentation. In particular, Gestalt psychology has uncovered a set of principles guiding the grouping process in the visual domain (Wertheimer, 1923; Koffka, 1935; Rock & Palmer, 1990). These principles work together to produce segmentation. We briefly summarize some of the most important principles (see also Rock and Palmer, 1990):

- **Proximity.** The closer the features lie to each other, the easier they are to be grouped into the same segment.
- **Similarity.** Features that have similar attributes, such as grayness, color, depth, or texture, tend to group together.
- **Common fate.** Features that have similar temporal behavior tend to group together. For instance, a group of features that move coherently (common motion) would form a single object. Notice that common fate may be regarded as one aspect of similarity. We list it separately to emphasize the importance of time as a separate dimension.
- **Connectedness.** A uniform, connected region, such as a spot, line, or more extended area, tends to form a single segment.
- **Good continuation.** A set of features that form a smooth and continuous curve tend to group together.
- **Prior knowledge.** If a set of features belong to the same familiar pattern, they tend to group together.

In computer vision algorithms for image segmentation, the result of segmentation can be represented in many ways. However, it is not a trivial task to represent the outcome of segmentation in a neural network. One proposal is naturally derived from the so-called neuron doctrine (Barlow, 1972), where neurons at higher brain areas are assumed to become more selective and eventually a single neuron represents each single object (the grandmother cell representation). Multiple objects in a visual scene would be represented by the coactivation of multiple units at some level of the nervous system. This representation faces major theoretical and neurobiological problems (von der Malsburg, 1981; Abeles, 1991; Singer, 1993). Another proposal relies on temporal correlation to encode the binding (Milner, 1974; von der Malsburg, 1981; Abeles, 1982). In particular, the correlation theory of von der Malsburg (1981) asserts that an object is represented by the temporal correlation of the firing activities of the scattered cells that encode different features of the object. Multiple objects are represented by differ-
ent correlated firing patterns that alternate in time, each corresponding to a single object.

Temporal correlation provides an elegant way to represent the result of segmentation. A special form of temporal correlation is oscillatory correlation, where the basic unit is a neural oscillator (see Terman & Wang, 1995; Wang & Terman, 1995a). However, this representation does not by itself reveal how segmentation is achieved using Gestalt grouping principles. Despite an extensive body of literature dealing with segmentation using temporal correlation (starting perhaps from von der Malsburg & Schneider, 1986), little progress has been made in building successful neural systems for image segmentation. There are two major challenges facing the oscillatory correlation theory. The first challenge is how to achieve fast synchronization within a population of locally coupled oscillators. Most of the models proposed for achieving phase synchrony rely on all-to-all connections (see Section 2 for more details). However, as Sporns, Tononi, and Edelman (1991) and Wang (1993a) pointed out, a network with full connections indiscriminately connects all the oscillators activated simultaneously by different objects because the network is dimensionless and loses critical information about geometry. The second challenge is how to achieve fast desynchronization among different groups of oscillators representing distinct objects. This is necessary in order to segment multiple objects simultaneously presented.

We have previously proposed a neural network framework to deal with the problem of image segmentation, called locally excitatory globally inhibitory oscillator networks (LEGION) (Wang & Terman, 1995a; Terman & Wang, 1995). Each oscillator is modeled as a standard relaxation oscillator. Local excitation is implemented by positive coupling between neighboring oscillators, and global inhibition is realized by a global inhibitor. LEGION exhibits the mechanism of selective gating, whereby oscillators stimulated by the same pattern tend to synchronize due to local excitation and oscillator groups stimulated by different patterns tend to desynchronize due to global inhibition (Wang & Terman, 1995a; Terman & Wang, 1995). We have proven that with the selective gating mechanism, LEGION rapidly achieves both synchronization within groups of oscillators that are stimulated by connected regions and desynchronization between different groups. In sum, LEGION provides an elegant solution to both challenges outlined above.

In this article, we study LEGION for segmenting real images. Before we demonstrate image segmentation, the original version of LEGION needs to be extended to handle images with many tiny (noisy) regions. One such example is shown in Figure 1, where three objects with a noisy background form a visual image. Without extension, LEGION would treat each region, no matter how small it is, as a separate segment. Thus, it would lead to many tiny fragments. We call this problem fragmentation. A more serious problem is that it is difficult to choose parameters so that LEGION is able to achieve more than several (5 to 10) segments (Terman & Wang, 1995). Noisy fragments may therefore compete with major image regions for becoming
segments, so that it may not be possible to extract the major segments from an image. The problem of fragmentation is solved by introducing a concept of lateral potential for each oscillator. The extended dynamics is fully analyzed (Wang & Terman, 1996), and the resulting LEGION network is applied to gray-level images and yields successful segmentation. A preliminary version of this work was presented in Wang and Terman (1995b).

In the next section we review prior work relevant to image segmentation and neural networks. In section 3, our model is described in detail. In section 4, computer simulations of the extended LEGION network are presented. Section 5 presents the segmentation results on real images. Further discussions concerning our approach are given in section 6.

2 Related Work

2.1 Image Segmentation Algorithms. Due to its critical importance for computer vision, image segmentation has been studied extensively. Many techniques have been invented (for reviews of the subject see Zucker, 1976;
Haralick, 1979; Haralick & Shapiro, 1985; Sarkar & Boyer, 1993b). There are three broad categories of algorithms: pixel classification, edge-based organization, and region-based segmentation. A simple classification technique is thresholding: A pixel is assigned a specific label if some measure of the pixel passes a certain threshold. This idea can be extended to a complex form, including multiple thresholds, which are determined by pixel histograms (Kohler, 1981). Edge-based techniques generally start with an edge-detection algorithm, which is followed by grouping edge elements into rectilinear or curvilinear lines. These lines are then grouped into boundaries that can be used to segment images into various regions (see, for example, Geman, Geman, Graffigne, & Dong, 1990; Sarkar & Boyer, 1993a; Foresti, Murino, Regazzoni, & Vernazza, 1994). Finally, region-based techniques operate directly on regions. A classical method is region growing and splitting (or split-and-merge; see Horowitz & Pavlidis, 1976; Zucker, 1976; Adams & Bischof, 1994), where iterative steps are taken to grow (or split) pixels into a connected region if all the pixels in the region satisfy some conditions. One of the apparent deficits with these algorithms is their iterative (serial) nature (Liou, Chiu, & Jain, 1991). There are some recent algorithms that are partially parallel (Liou, Chiu, & Jain, 1991; Mohan & Nevatia, 1992; Manjunath & Chellappa, 1993).

Most of these techniques rely on domain-specific heuristics to perform segmentation, and no unified computational framework exists to explain the general phenomenon of scene segmentation (Haralick & Shapiro, 1985). The problem of scene segmentation is computationally hard (Gurari & Wechsler, 1982) and largely regarded as unsolved.

2.2 Neural Network Efforts. Neural networks have proven to be a successful approach to pattern recognition (Schalkoff, 1992; Wang, 1993b). Unfortunately, little work has been devoted to scene segmentation, which is generally regarded as part of preprocessing (often meaning manual segmentation). Scene segmentation is a particularly challenging task for neural networks, partly because traditional neural networks lack the representational power for encoding multiple objects simultaneously. Interesting schemes of segmentation based on learning have been proposed (Sejnowski & Hinton, 1987; Mozer, Zemel, Behrmann, & Williams, 1992). Grossberg and Wyse (1991) proposed a model for segmentation based on the contour detection model of Grossberg and Mingolla (1985; see Gove, Grossberg, & Mingolla, in press, for a computer simulation). However, all of these methods were tested on only small synthetic images, and it is not clear how they can be extended to handle real images. Also, Kohonen’s self-organizing maps have been used for segmentation based on pixel classification (Kohonen, 1995; Koh, Suk, & Bhandarkar, 1995). A primary drawback of these methods is that the number of segments (objects) is assumed to be known a priori.

Because temporal (oscillatory) correlation offers an elegant way of representing multiple objects in neural networks (von der Malsburg & Schnei-
most of the neural network efforts on image segmentation have centered around this theme. In particular, the discovery of synchronous oscillations in the visual cortex has triggered much interest in exploring oscillatory correlation to solve the problems of segmentation and figure-ground segregation. One type of model uses all-to-all connections to reach synchronization (Wang, Buhmann, & Malsburg, 1990; Sompolinsky, Golomb, & Kleinfeld, 1991; von der Malsburg and Buhmann, 1992). As explained in section 1, these models cannot extend very far in solving the segmentation problem because fundamental information concerning the geometry among sensory features is lost. Another type of model uses lateral connections to reach synchrony (Sporns, et al., 1991; Murata & Shimizu, 1993; Schillen & König, 1994). Unfortunately, it is unclear to what extent these oscillator networks can synchronize on the basis of local connectivity since no analysis is given and only simulation results on small networks are provided. Moreover, recent insights into the contrasting behavior between sinusoidal and relaxation oscillators make it clear that sinusoid-typed oscillators, which encompass most of the oscillator models used, have severe limitations to support fast synchronization (Wang, 1995; Terman & Wang, 1995; Somers & Kopell, 1995). In fact, in all of the above models, nothing close to a real image has ever been used for testing these models.

3 Model Description

The building block of LEGION, a single oscillator $i$, is defined as a feedback loop between an excitatory unit $x_i$ and an inhibitory unit $y_i$, whose time derivatives are defined as

$$
\begin{align}
    x'_i &= 3x_i - x_i^3 + 2 - y_i + I_i H(p_i + \exp(-\alpha t) - \theta) + S_i + \rho \\
    y'_i &= \varepsilon(\gamma(1 + \tanh(x_i/\beta)) - y_i).
\end{align}
$$

(3.1a)

Here $H$ stands for the Heaviside step function, which is defined as $H(v) = 1$ if $v \geq 0$ and $H(v) = 0$ if $v < 0$. $I_i$ represent external stimulation, which is assumed to be applied from time 0 on, and $S_i$ denotes the coupling from other oscillators in the network. $\rho$ denotes the amplitude of gaussian noise, the mean of which is set to $-\rho$. The negative mean is used to reduce the chance of self-generating oscillations, which will become clear in the next paragraph. The noise term is introduced for two purposes. The first one is obvious: to test the robustness of the system. The second one, perhaps more important, is to play an active role in separating different input patterns (for more discussion, see Terman & Wang, 1995).

The parameter $\varepsilon$ is a small, positive number. Hence equation 3.1, without any coupling or noise and with constant stimulation, corresponds to a standard relaxation oscillator. The $x$-nullcline of equation 3.1 is a cubic curve, while the $y$-nullcline is a sigmoid function. If $I > 0$ and $H = 1$, these curves
intersect along the middle branch of the cubic when $\beta$ is small. In this case, we call the oscillator enabled (see Figure 2A). It produces a stable periodic orbit, which alternates between silent and active phases of near steady-state behavior. As shown in Figure 2A, the silent and the active phases correspond to the left $\mathcal{L}$ and the right $\mathcal{R}$ branches of the cubic, respectively. The transitions between the two phases occur rapidly (and are thus referred to as jumping). Notice that the trajectory of the oscillator in phase space jumps between the two branches and then follows closely the branches, because the small $\varepsilon$ induces very different time scales for $x$- and $y$-dynamics. The parameter $\gamma$ is used to control the ratio of the times that the solution spends in these two phases. For a larger value of $\gamma$, the solution spends a shorter time in the active phase. If $I \leq 0$ and $H = 1$, the nullclines of equation 3.1 intersect at a stable fixed point along the left branch of the cubic (see Fig. 2B). In this case, equation 3.1 produces no periodic orbit, and the oscillator is referred to as excitable, indicating that the oscillator has not yet been but can be excited by stimulation. An excitable oscillator may become oscillatory if it receives, through the term $S$, large enough coupling from other oscillators. Because of this dependency on external stimulation, the oscillations are stimulus dependent. We say that the oscillator is stimulated if $I > 0$, and unstimulated if $I \leq 0$. The parameter $\beta$ specifies the steepness of the sigmoid function and is chosen to be small. The oscillator model (see equation 3.1) may be interpreted as a model for the spiking behavior of a single neuron, the envelope of a bursting neuron, or a mean field approximation to a network of excitatory and inhibitory binary neurons.

The primary difference between equation 3.1 and the model in Terman and Wang (1995) is the introduction of the Heaviside function in which $\alpha > 0$ and $0 < \theta < 1$. The parameter $\alpha$ is chosen to be on the same order of magnitude as $\varepsilon$ so that the exponential function decays on a slow time scale. It is the Heaviside term that allows the network to distinguish between major blocks and noisy fragments. The basic idea is that a major block must contain at least one oscillator, denoted as a leader, which lies in the center of a large, homogeneous region. This oscillator will be able to receive large lateral excitation from its neighborhood. A noisy fragment does not contain such an oscillator. The variable $p_i$ in equation 3.1a determines whether an oscillator is a leader. It is referred to as the lateral potential of the oscillator $i$, and it satisfies the differential equation:

$$p_i' = \lambda (1 - p_i) H \left[ \sum_{k \in N(i)} T_{ik} H(x_k - \theta_x) - \theta_p \right] - \mu p_i.$$  

(3.2)

Here $\lambda > 0$, $T_{ik}$ is the permanent connection weight (explained later) from oscillator $k$ to $i$, and $N(i)$ is called the neighborhood of $i$. If the weighted sum oscillator $i$ receives from $N(i)$ exceeds the threshold $\theta_p$, $p_i$ approaches 1. If this weighted sum is below $\theta_p$, $p_i$ relaxes to 0 on a time scale determined by $\mu$, which is chosen to be on the same order as $\varepsilon$ resulting in a slow time scale.
Figure 2: Nullclines and orbits of a single oscillator. (A) If $I > 0$ and $H = 1$, the oscillator is enabled. The periodic orbit is shown with a bold curve, and its direction of motion is indicated by the arrowheads. The left and the right branches of the $x$-nullcline are labeled $L$ and $R$, respectively. $LK$ and $RK$ indicate the left and the right knees of the cubic, respectively. (B) If $I \leq 0$ and $H = 1$, the oscillator is excitable. The fixed point $P_I$ on the left branch of the cubic is asymptotically stable.
Figure 3: Architecture of a two-dimensional LEGION network with four nearest-neighbor coupling. An oscillator is indicated by an open circle, and the global inhibitor is indicated by the filled circle.

It follows that $p_i$ can exceed the threshold $\theta$ in equation 3.1a only if $i$ is able to receive a large enough lateral excitation from its neighborhood. In order to develop a high potential, it is not sufficient that a large number of neighbors of $i$ are oscillatory. They must also have a certain degree of synchrony in their oscillations. In particular, they all must exceed the threshold $\theta_x$ at the same time in their oscillations.

The purpose of introducing the lateral potential is that an oscillator with a high potential can lead the activation of an oscillator block corresponding to an object. Although a high-potential oscillator need not be stimulated, it must be stimulated in order to play the role of leading an oscillator block; otherwise, the oscillator will not oscillate at all. Thus, we require that a leader be always stimulated. More formally, an oscillator $i$ is defined as a leader if $p_i \geq \theta$ and $i$ is stimulated. The lateral potential of every oscillator is initialized to zero.

The network we study for image segmentation is two-dimensional. Figure 3 shows the simplest case of permanent connectivity, where an oscillator is connected only with its four immediate neighbors except on the boundaries where no wraparound is used. Such connectivity forms a two-dimensional grid. In general, however, $N(i)$ should be larger, and the permanent connection weights should take on the form of a gaussian distribution with their distance.

The coupling term $S_i$ in equation 3.1 is given by

$$S_i = \sum_{k \in N(i)} W_k H(x_k - \theta_x) - W_z H(z - \theta_z).$$

(3.3)
where $W_{ik}$ is the dynamic connection weight from $k$ to $i$. The neighborhood of the above summation is chosen to be the same as equation 3.2. In some situations, however, they should be chosen differently to achieve good results, and an alternative definition with two different neighborhoods is given elsewhere (Wang & Terman, 1996).

Now let us explain permanent and dynamic connection weights. To facilitate synchrony and desynchrony, we assume that there are two kinds of synaptic weights (links) between two oscillators following von der Malsburg who argued for its neurobiological plausibility (von der Malsburg, 1981; von der Malsburg & Schneider, 1986; see also Crick, 1984). The permanent weight, or $T_{ik}$, embodies the hardwired structure of a network. On the other hand, the dynamic weight, or $W_{ik}$, rapidly changes. $W_{ik}$ is formed on the basis of $T_{ik}$ according to the mechanism of dynamic normalization (Wang, 1995). Dynamic normalization was previously defined as a two-step procedure: first update dynamic links and then normalization (Wang, 1995; Terman & Wang, 1995). There are different ways to realize such normalization. In the following, we give one way to implement dynamic normalization in differential equations,

$$u_i' = \eta(1 - u_i)I_i - \nu u_i$$

$$W'_{ik} = W_{Tik}u_iu_k - W_{ik}\sum_{j \in N(i)} T_{ij}u_iu_j.$$  

(3.4a)

(3.4b)

The function $u_i$ measures whether oscillator $i$ is stimulated, and it is initialized to 0. The parameter $\eta$ determines the rate of updating $u_i$. When $I_i > 0$, $u_i \to 1$ quickly because we choose $\eta \gg \nu$ (see below); otherwise when $I_i = 0$, $u_i = 0$. For this equation we assume $I_i = 0$ if oscillator $i$ is unstimulated (otherwise it is easy to enforce this by applying a step function on $I_i$). The parameter $\nu$ is chosen to be on the same order as $\varepsilon$, so that $u_i$ slowly relaxes back to 0 after the external stimulus is withdrawn.

We assume that $W_{ik}$ are initialized to 0 for all $i$ and $k$. It is easy to see that if oscillator $i$ is unstimulated, $W_{ik}$ remains to be 0 for all $k$, and if oscillator $k$ is unstimulated $W_{ik} = 0$ for all $i$. Otherwise, if $u_i = 1$ and $u_k = 1$ for at least one $k \in N(i)$, then at equilibrium,

$$W_{ik} = \frac{W_{Tik}u_iu_k}{\sum_{j \in N(i)} T_{ij}u_iu_j} \quad \text{and} \quad \sum_{k \in N(i)} W_{ik} = W_T.$$  

Thus the total dynamic weights converging to a single oscillator equals $W_T$, which gives the desired normalization. Notice that dynamic weights, not permanent weights, participate in determining $S_i$ (see equation 3.3). Moreover, $W_{ik}$ can be properly set up in one step at the beginning based on external stimulation, which should be useful for engineering applications.

Weight normalization is not a necessary condition for the selective gating mechanism to work. This conclusion has been established previously.
(Terman & Wang, 1995). With normalized weights, however, the quality of synchronization within each oscillator block is better (Terman & Wang, 1995).

In equation 3.3, $W_z$ is the weight of inhibition from the global inhibitor $z$, whose activity, also denoted by $z$, is defined as

$$z' = \phi(\sigma_\infty - z), \tag{3.5}$$

where $\sigma_\infty = 1$ if $x_i \geq \theta_{zx}$ for at least one oscillator $i$, and $\sigma_\infty = 0$ otherwise. Hence $\theta_{zx}$ represents another threshold, and it is chosen so that only an oscillator jumping to the active phase can trigger the global inhibitor. If $\sigma_\infty$ equals 1, $z \rightarrow 1$. The parameter $\phi$ represents the rate at which the inhibitor reacts to the stimulation from the oscillator network.

The introduction of a lateral potential provides a solution to the problem of fragmentation. There is an initial period when the term $\exp(\alpha t)$ exceeds the threshold $\theta$. During this period, every stimulated oscillator is enabled. This allows the leaders to receive sufficient lateral excitation so that they can achieve a high potential. After this initial period, the only oscillators that can jump up without stimulation from other oscillators are the leaders. When a leader jumps up, it spreads its activity to other oscillators within its own block, so they also can jump up. These oscillators are referred to as followers. Oscillators not in this block are prevented from jumping up because of the global inhibitor. The oscillators that belong to the noisy fragments will not be able to jump up beyond the initial period because these oscillators will not be able to develop a sufficiently high potential by themselves and they cannot be recruited by leaders. These oscillators are referred to as loners. In order to be oscillatory beyond the initial time period, an oscillator must be either a leader or a follower. This indicates that the oscillator is not part of a noisy fragment, because noisy fragments in an image tend to be small and isolated (see Figure 1). The collection of all noisy regions whose corresponding oscillators are loners is called the background, which is not a uniform region and generally discontinuous.

We have proven a number of rigorous results concerning the system equations 3.1 to 3.5. Our main result implies that the loners will no longer be able to oscillate after an initial time period. Moreover, the asymptotic behavior of a leader or a follower is precisely the same as the network obtained by simply removing all the loners. Together with the results in Terman and Wang (1995), this implies that after a number of oscillation cycles, a block of oscillators corresponding to a single major image region will oscillate in synchrony, while any two oscillator blocks corresponding to two major regions will desynchronize from each other. Also, the number of cycles required for full segmentation is no greater than the number of major regions plus one. The details of the analysis are given in Wang and Terman (1996).
The analysis in Wang and Terman (1996) is constructive in the sense that it leads to precise estimates that the parameters must satisfy. It shows that the results hold for a robust range of parameter values. Moreover, the analysis does not depend on the precise form of nonlinear functions in equation 3.1. The specific cubic and sigmoid functions (see Figure 2) are used because of their simplicity. In addition to the parameter description given earlier, we require that $0 < \theta < 1$, and $\alpha$ be chosen so that all the stimulated oscillators remain enabled for the first cycle, but only leaders remain enabled during the second cycle. In equation 3.2, we simply require that $\lambda$ be on the same order of magnitude as 1, and $0 < \theta_p < 1$. The parameter $\eta$ in equation 3.4 simply needs to be on the same order of 1.

There are alternative ways of defining the model without affecting its essential dynamics (Wang & Terman, 1996). In particular, we have given a definition where dynamic normalization of connection strengths in equation 3.4 is not needed, but the quality of synchrony within each block and the flexibility for choosing parameters seem somewhat lessened.

4 Computer Simulation

To illustrate how the LEGION network is used for image segmentation while eliminating fragmentation, we have simulated a $50 \times 50$ grid of oscillators with a global inhibitor as defined by equations 3.1 through 3.5. We map the three objects (designated as the sun, a tree, and a mountain) in Figure 1, and then add 20% noise so that each uncovered square has a 20% chance of being covered (stimulated). The resulting image is shown in Figure 4A. In the simulation, $N(i)$ is the four nearest neighbors without boundary wraparound. For all the stimulated oscillators $I = 0.2$, while for the others $I = 0$. Notice that if oscillator $i$ is unstimulated, $W_{ik} = W_{ki} = 0$ for all $k$, and $I_i$ does not need to be negative to prevent $i$ from oscillating. The amplitude $\rho$ of the gaussian noise was set to 0.02. This represents a 10 percent noise level compared to the external stimulation. We observed during the simulations that noise facilitated the process of desynchronization.

The differential equations equations (3.1–3.5) were solved using both a fourth-order Runge-Kutta method and the adaptive grid ordinary differential equations solver LSODE. Permanent connections between any two neighboring oscillators were set to 2.0, and for total dynamic connections (see equation 3.4b), $W_T = 8.0$. Dynamic weights $W_{ik}$ were set up at the beginning according to equation 3.4. The following values for the other parameters in equations 3.1 through 3.5 were used: $\epsilon = 0.02$, $\alpha = 0.005$, $\beta = 0.1$, $\gamma = 6.5$, $\theta = 0.9$, $\lambda = 0.1$, $\theta_{z} = -0.5$, $\theta_p = 7.0$, $W_z = 1.5$, $\eta = 1.0$, $\mu = \nu = 0.01$, $\phi = 3.0$, and $\theta_{zx} = \theta_{xz} = 0.1$. The value of $\theta_p$ was chosen so that in order to achieve a high potential, an oscillator must have all four neighbors active. The simulation results were robust to considerable changes in the parameter values. Figures 4B–E shows the instantaneous activity (snapshot) of the network at various stages of dynamic evolution. The diameter
Figure 4: (A) An image composed of three patterns on a noisy background. The image is mapped to a 50 × 50 LEGION network. Each square corresponds to an oscillator. If a square is entirely covered, the corresponding oscillator receives external input; otherwise, the oscillator receives no external input. In the figure, B–E correspond to the case with the inclusion of the lateral potential, whereas F–J correspond to the case without the lateral potential. (B) A snapshot at the beginning of dynamic evolution. (C–E) Snapshots subsequently taken shortly after B. (F) A snapshot at the beginning of dynamic evolution for the case without the lateral potential. (G–J) Snapshots subsequently taken shortly after F.
of each black circle represents the $x$ activity of the corresponding oscillator. Specifically, if the range of $x$ values of all the oscillators is given by $x_{\text{min}}$ and $x_{\text{max}}$, then the diameter of the black circle corresponding to one oscillator is set to be proportional to $(x - x_{\text{min}})/(x_{\text{max}} - x_{\text{min}})$.

Figure 4B shows the snapshot at the beginning of the dynamic evolution. This is included to illustrate the random initial conditions. Figure 4C shows a snapshot shortly after Figure 4B. The effect of synchrony and desynchrony is clear: All the stimulated oscillators that belong to or are the neighbors of the sun are entrained and have large activities (in the active phase). At the same time, the oscillators stimulated by the rest of the image have very small activities (in the silent phase). Thus the noisy sun is segmented from the rest of the image. A short time later, as shown in Figure 4D, the oscillators in the group representing the noisy tree reach their active phase and are separated from the rest of the image. Figure 4E shows another snapshot, when the noisy mountain has its turn to be activated and separate from the rest of the input. This successive “pop-out” of the segments continues in a stable periodic fashion until the input image is withdrawn. To illustrate the entire segmentation process, Figure 5 shows the temporal evolution of every stimulated oscillator. The activities of the oscillators stimulated by each noisy object are combined together as one trace in the figure, and so are for the background. Since the oscillators receiving no external stimulation remain excitable and unable to oscillate throughout the simulation process, they are excluded from the display. The three upper traces represent the activities of the three oscillator blocks, and the fourth represents the background, consisting of all the scattered dots. Because of low potentials, these oscillators quickly become excitable even though they are enabled at the beginning. The bottom trace represents the activity of the global inhibitor. The synchrony within each block and desynchrony between different blocks are clearly shown after three cycles.

To illustrate the role of the lateral potential, the same network with the same input (see Figure 4A) and the same initial condition has been simulated without the lateral potential. In this case, the Heaviside function in equation 3.1a is always 1 for every oscillator. With the random initial condition shown in Figure 4F, the network reaches a stable oscillatory behavior with four segments after fewer than three cycles. The four segments are shown in Figures 4G–J. Without the lateral potential, the network cannot distinguish major image regions from noisy fragments and separate major regions apart.

With a fixed set of parameters, the dynamical system of LEGION can segment only a limited number of patterns. This number depends to a large extent on the ratio of the times that a single oscillator spends in the silent and active phases. Let us refer to this limit as the segmentation capacity of LEGION. In the above simulation, the number of the major blocks to be segmented is within the segmentation capacity. What happens if this number exceeds the segmentation capacity? From the analysis in Wang and Terman
Figure 5: Temporal evolution of every stimulated oscillator. The upper three traces show the combined x activities of the three oscillator blocks representing the three corresponding patterns indicated by their respective labels. The fourth trace shows the temporal activities of the loners, and the bottom trace shows the activity of the global inhibitor. The ordinates indicate the normalized activity of an oscillator or the inhibitor. The simulation took 9000 integration steps.

(1996), we know that the system will separate the entire image into as many segments as the capacity allows, where each segment may correspond to one major block (called a simple segment) or a number of major blocks (called a congregate segment). To illustrate this point, we show the following simulation, where we present to a 30 × 30 LEGION network with an arbitrary image containing nine binary patterns, which together form the phrase OHIO STATE as shown in Figure 6A. We then add 10 percent random noise to the input in a similar way as in Figure 4, resulting in Figure 6B. We use the same parameter values as in the simulations presented in Figure 4, except that γ = 8.0. For this set of parameters, our earlier experiments showed that the system’s segmentation capacity is less than 9. The simulation results are presented in Figures 6C–H. Shortly after the start of system evolution, the LEGION network segmented the input of Figure 6B into five segments, shown in Figures 6D–H, respectively. Among these five segments, three are simple segments (6D, 6E, and 6H) and two are congregate segments (6F and 6G). Besides Figure 6, many other simulations have been performed for the input of Figure 6B with different random initial conditions, and the results are comparable with Figure 6. There are different ways, however, that the
system separates the nine noisy patterns into five segments. For this particular set of parameters, the segmentation capacity of the LEGION network is five. In fact, we have not seen a single simulation trial where more than five segments are produced. This important property of the system—it naturally exhibits a segmentation capacity—is in good accord with the well-known psychological principle that there are fundamental limits on the number of simultaneously perceived objects.
5 Real Images

LEGION can segment gray-level images in a way similar to segmenting binary images. For a given image, a LEGION network of the same size as the image with a global inhibitor is used to perform segmentation. Each pixel of the image corresponds to an oscillator of the network, and we assume that every oscillator is stimulated when the image is applied to the network. The main difference between gray-level and binary images lies in how to set up connections. For gray-level images, the coupling strength between two neighboring oscillators is determined by the similarity of two corresponding pixels. This simple way of setting up the coupling strength addresses only the grouping principles of proximity, connectedness, and similarity (cf. section 1).

5.1 Algorithm. To segment real images with large numbers of pixels involves integrating a large number of the differential equations (equations 3.1–3.5). To reduce numerical computations on a serial computer, an algorithm is extracted from these equations. The algorithm follows major steps in the numerical simulation of the equations, and it exhibits the essential properties of relaxation oscillators, such as two time scales (fast and slow) and the properties of synchrony and desynchrony in a population of oscillators. Such extraction is quite straightforward because, in a relaxation oscillator network, much of the dynamics takes place when oscillators are jumping up or jumping down. Besides, the algorithm overcomes the segmentation capacity, which may be desired in some applications. More specifically, the following approximations have been made:

1. When no oscillator is in the active phase (see Figure 2), the leader closest to the jumping point (left knee) among all enabled oscillators is selected to jump up to the active phase.

2. An oscillator takes one time step to jump up to the active phase if the net input it receives from neighboring oscillators and the global inhibitor is positive.

3. The alternation between the active phase and the silent phase of a single oscillator takes one time step only.

4. All of the oscillators in the active phase jump down if no more oscillators can jump up. This situation occurs when the oscillators stimulated by the same pattern have all jumped up.

LEGION algorithm. Only the $x$ value of oscillator $i, x_i$, is used in the algorithm. $N(i)$ is assumed to be the eight nearest neighbors of $i$ without wraparound. $LK_x, RK_x, LC_x$ represent the $x$ values of three of four corner points of a typical limit cycle (see Figure 2A), where $LC$ denotes the (upper) left corner of the limit cycle. By straightforward calculations, we obtain
$L_{K_x} = -1, L_{C_x} = -2, R_{K_x} = 1$. In the algorithm, $I_i$ indicates the value of pixel $i$, and $I_M$ indicates the maximum possible pixel value.

1. Initialize
   1.1 Set $z(0) = 0$;
   1.2 Form effective connections
      \[ W_{ij} = I_M/(1 + |I_i - I_k|), \quad k \in N(i) \]
   1.3 Find leaders
      \[ p_i = H \left[ \sum_{k \in N(i)} W_{ik} - \theta_p \right] \]
   1.4 Place all the oscillators randomly on the left branch. Namely, $x_i(0)$ takes a random value between $L_{C_x}$ and $L_{K_x}$.

2. Find one oscillator $j$ so that (1) $x_j(t) \geq x_k(t)$, where $k$ is currently on the left branch and $P_k = 1$; (2) $p_j = 1$. Then:
   \[
   \begin{align*}
   x_j(t+1) &= R_{K_x} \quad z(t+1) = 1 \quad \text{[jump up]} \\
   x_k(t+1) &= x_k(t) + (L_{K_x} - x_j(t)) \quad \text{for } k \neq j \text{ and } P_k = 1.
   \end{align*}
   \]
   In this step, the leader on the left branch, which is closest to the left knee, is selected. This leader jumps up to the right branch, and all the other leaders move toward $L_{K_x}$.

3. Iterate until stop
   If $(x_i(t) = R_{K_x} \text{ and } z(t) > z(t-1))$
      \[ x_i(t+1) = x_i(t) \quad \text{[stay on the right branch]} \]
   else if $(x_i(t) = R_{K_x} \text{ and } z(t) \leq z(t-1))$
      \[ x_i(t) = L_{C_x} ; z(t+1) = z(t) - 1 \quad \text{[jump down]} \]
   If $(z(t+1) = 0)$ go to step 2
   else
      \[ S_i(t+1) = \sum_{k \in N(i)} W_{ik}H(x_k(t)) - W_zH(z(t) - 0.5) \]
      If $(S_i(t+1) > 0)$
         \[ x_i(t+1) = R_{K_x} ; z(t+1) = z(t) + 1 \quad \text{[jump up]} \]
      else
         \[ x_i(t+1) = x_i(t) \quad \text{[stay on the left branch]} \]

In a comparison of this algorithm with the dynamical system of equations 3.1–3.5, one can find the following simplifications.

1. The dynamic weight $W_{ij}$ is directly set to $W_{ij} = I_M/(1 + |I_i - I_k|)$. The intuitive reason for this choice of weights is that the more pixel
i and pixel j are similar to each other, the stronger is the connection between the two corresponding oscillators. It is worth noting that the algorithm does not compute normalized weights. As mentioned in section 3, selective gating can still take place with this weight setting, even though the weights are not normalized.

2. The leaders are chosen during initialization. According to the dynamics described in section 3, lateral potentials, and thus leaders, are determined during a few initial cycles of oscillatory dynamics. Since every oscillator is stimulated and $W_{ij}$ is set at the beginning, it can be precisely predicted at the beginning which oscillators will become leaders. Thus, to save computational time, the leaders are determined in the initialization step. It should be clear that the number of leaders determined in this step does not correspond to the number of resulting segments; a major image region (segment) may generate many leaders.

There are two critical parameters in the algorithm: $W_z$ and $\theta_p$, where the former is the strength of global inhibition and the latter is the threshold for forming high potentials (leaders). For $W_z$, higher values make the algorithm more difficult to group pixels into regions. Thus, in order for a region to be grouped together, the algorithm demands a higher degree of homogeneity within the region. Generally, given a gray-level image, higher $W_z$ leads to more and smaller regions. For $\theta_p$, higher values make the algorithm more difficult to develop leaders. Thus fewer leaders will be developed, and fewer regions result from the algorithm. On the other hand, regions produced with a higher $\theta_p$ tend to be more homogeneous. For image segmentation applications, it suffices to stop the algorithm when every leader has jumped up once. (See Wang & Terman, 1996, for some discussions on the algorithm.)

5.2 Segmenting Real Images.

5.2.1 Summation versus Maximation: An Aerial Image. The first image the algorithm is tested on is an aerial image, called Lake, which is shown in Figure 7A. As in the following images to be used, this is a typical gray-level image, where each pixel is an 8-bit number ranging from 0 to 255 (also called intensity), and pixels with higher values appear brighter. The image has $160 \times 160$ pixels and is presented to a LEGION network of $160 \times 160$ oscillators. For this simulation, $W_z = 40$ and $\theta_p = 1200$. Quickly after the image is presented, the algorithm produces different segments at different time steps. Figures 7B–G display the first six segments that have been produced sequentially, where a black pixel corresponds to an oscillator in the active phase and a blank pixel corresponds to an oscillator in the silent phase. Each segment in the figure corresponds to a meaningful region in the original image: A segment is a lake, a field, or a parkway. The region in Figure 7B corresponds to a lake. The region in Figure 7C corresponds to the
Figure 7: (A) A gray-level image consisting of 160 × 160 pixels (courtesy of K. Boyer). (B–G) Segments popped out subsequently from the network shortly after the LEGION algorithm is executed.

main lake, except for the lower-left part and on the right side, where the lake region extends to nonlake parts. The parkway segment in Figure 7G picks up a partial parkway network in the original image. The other segments match well with the fields of the image.

The entire image is separated into 16 regions and a background. To simplify the display, we put all the segments and the background together into one figure, using gray levels to indicate the phases of oscillator blocks. Such a display is called a gray map. The gray map of the results of this simulation is shown in Figure 8A, where the background is shown by the black scattered areas. Generally, the background corresponds to parts of the im-
Figure 8: (A) A gray map showing the result of segmenting Figure 7A. The algorithm produces 16 segments plus a background. (B) The result of another segmentation using maximization to compute $S$. The algorithm produces 17 segments plus a background. (C) The result of another segmentation similar to B but with a different value for $\theta_p$. The system produces 23 segments plus a background. The algorithm was run for 1000 steps for every case.

Image Segmentation
the oscillator jumps to the active phase (see also equation 3.3). Another reasonable way of grouping is to replace summation by maximization when computing $S_i$:

$$S_i = \max_{k \in \mathcal{N}(i)} \{W_k H(x_k - \theta_x) \} - W_z H(z - \theta_z).$$

(5.1)

Intuitively, the maximum operation concentrates on the relation of $i$ with the oscillator in $\mathcal{N}(i)$ that has the strongest coupling with $i$ but omits the relation between $i$ and $\mathcal{N}(i)$ as a whole. Thus, grouping by maximization emphasizes pairwise pixel relations, whereas grouping by summation emphasizes pixel relations in a local field.

By using equation 5.1 in the LEGION algorithm, the Lake image is segmented again. In this simulation, $W_z = 20$ and $\theta_p = 1200$. Figure 8B shows the result of segmentation by a gray map. The entire image is segmented into 17 regions and a background, which is indicated by black areas. Each segment corresponds well with a relatively homogeneous region in the image. Interestingly, except for the parkway region in the lower part of the image, every region in Figure 8B has a corresponding one in Figure 8A. A comparison between the two figures reveals the difference between summation and maximization in segmentation. A closer comparison, however, indicates that the maximum scheme yields a little more faithful regions. On the other hand, regions in Figure 8A appear smoother and have fewer “black holes”—parts of the background. The smoothing effect of summation is generally positive, but it may lose important details. For example, the sizable hole inside the main lake region of Figure 8B corresponds to an island in the original image, which is neglected in the main lake region of Figure 8A. Another distinction is that grouping in the maximum scheme is symmetrical in the sense that if pixel $a$ can recruit pixel $b$, then $b$ can recruit $a$ as well. This is because effective weights are symmetrical, namely, $W_{ij} = W_{ji}$ (see the algorithm). Because the maximum scheme appears to produce better results, it will be used in all of the following simulations.

To show the effects of parameters, we reduce the value of $\theta_p$ from 1200 in Figure 8B to 1000. As a result, more regions are segmented, as shown in Figure 8C, where 23 segments plus a background are produced by the algorithm. Compared with Figure 8B, the notable new segments include an open-theater-like region to the left of the main lake and its nearby field region. The Lake image has been used by Sarkar and Boyer (1993a), whose approach is edge based. Readers are encouraged to compare our results with theirs.

5.2.2 MRI Images. The next image to test our algorithm is an MRI (magnetic resonance imaging) image of a human head (see Figure 9A). MRI images constitute a large class of medical images, and their automatic processing is of great practical value. This particular image, which we denote as Brain-1, is a midsagittal section, consisting of $257 \times 257$ pixels. Salient regions
of this picture are the cerebral cortex, the cerebellum, the brainstem, the corpus callosum, the fornix (the bright stripe below the corpus callosum), the septum pellucidum (the region surrounded by the corpus callosum and the fornix), the extracranial soft tissue (the bright stripe on top of the head), the bone marrow (scattered stripes under the extracranial tissue), and several other structures (for the nomenclature see Kandel, Schwartz, & Jessell, 1991, p. 318). For this image, a LEGION network of 257 × 257 oscillators is used, and \( W_z = 25 \) and \( \theta_p = 800 \). Figure 9B shows the result of one simulation by a gray map. The Brain-1 image is segmented into 21 regions plus a background, indicated by the black areas. Of particular interest are two parts of the brain: (1) the upper part and (2) the brainstem with part of the spinal cord (“brainstem” for short), parts of the extracranial tissue, and parts of the bone marrow. Other interesting segments are the neck part, the chin part, the nose part, and the vertebral segment.

Although it is useful to treat the brain as a whole in some circumstances, it may also be desirable to segment the brain into more detailed structures. To achieve this in LEGION, one can increase \( W_z \). But in order not to produce too many regions, \( \theta_p \) should usually be increased as \( W_z \) increases. To show the combined effects, Figure 9C displays the result of another run with \( W_z = 40 \) and \( \theta_p = 1000 \), where Brain-1 is segmented to 25 regions plus a background. Now the upper part of the brain is further segmented into the cerebral cortex, the cerebellum, the callosum-fornix region, and its surrounding septum. Because of the higher \( W_z \), regions in Figure 9C tend to contain more background (compare, for example, the two brainstem regions). However, in the cortex segment in Figure 9C, the noisy stripes actually have physical meanings; they tend to match with various fissures on the cerebral cortex. With the higher \( \theta_p \), some segments in Figure 9B cannot generate any leaders and thus join the background.

The final segmentation uses another MRI image of a human head, shown in Figure 10A. This image is denoted as Brain-2, consisting of 257 × 257 gray-level pixels. Brain-2 is a sagittal section through one eye. Salient regions of this picture are the cortex, the cerebellum, the lateral ventricle (the black hole within the cortex), the eye, the sinus (the black hole below the eye), the extracranial soft tissue, and the bone marrow. A LEGION network with 257 × 257 oscillators is used for the segmentation task. In the first simulation, \( W_z = 20 \) and \( \theta_p = 800 \); Figure 10B shows the result by a gray map. Brain-2 is segmented into 17 regions plus a background, indicated by black scattered areas. One can see from the figure that the entire brain forms a single segment. Other significant segments are the eye, the sinus, parts of the bone marrow, and parts of the extracranial tissue. The lateral ventricle is put into the background.

In order to generate finer structures, \( W_z \) is raised to 35 in the second simulation. As in Figure 9, \( \theta_p \) is increased to 1000. Figure 10C shows the result of this simulation, where Brain-2 is segmented into 13 regions plus a background. As expected, the segments in Fig. 10B become further segmented...
Figure 9: (A) A gray-level image consisting of $257 \times 257$ pixels (courtesy of N. Shareef). (B) A gray map showing the result of segmenting the image by a $257 \times 257$ LEGION network. The system produces 21 segments plus a background. (C) The result of another segmentation with different values of $W_c$ and $\theta_p$. The system produces 25 segments plus a background. The algorithm was run for 1200 steps in both B and C.

or shrunk, and the background becomes more extensive. Worth mentioning is that the brain segment in Figure 10B is segmented into three segments: one corresponding to the cortex and the other two corresponding to the cerebellum. Due to the increase of $\theta_p$, the segments corresponding to the extracranial tissue and the marrow in Figure 10B disappear in Figure 10C.
In the segmentation experiments of this section, our goal was to illustrate the LEGION mechanism and the potential effectiveness of derived algorithms. We did not attempt to produce best possible results by fine tuning of parameters. One can easily tell this by the simple rule of setting $W_{ij}$, the simple choice of $N(i)$, and the values of $W_z$ and $\theta_p$ that have been
used in the simulations. Therefore, better results can be expected by using more sophisticated schemes of choosing these parameters. Indeed, a more elaborate version of the LEGION algorithm has been applied to segment three-dimensional MRI and CT (computerized tomography) images, and good segmentation results have been obtained (Shareef et al., in preparation).

6 Discussion

6.1 Further Remarks on LEGION Computation. In the simulations of section 5, \( N(i) \) is set to the eight nearest neighbors of \( i \). Larger \( N(i) \)'s entail more computations for determining leaders. But larger \( N(i) \)'s have more flexibility in specifying the conditions for creating leaders, which tends to produce better results. Also, the order of pop-out of different segments is currently random. In some situations, however, it might be useful to influence the order of pop-out by some criteria, such as the size of each region. LEGION may incorporate different criteria by ordering leaders accordingly, for example, by using the global inhibitor in different ways.

The main difference between our approach to image segmentation and other segmentation algorithms reviewed in section 2 is that ours is neurocomputational, building on the strengths and the constraints of neural computation. Our approach relies on emergent behavior of LEGION to embody the computation involved in image segmentation. Methodologically, the computation is performed by a population of active agents—oscillators, in this case—that are driven by pixels, whereas in a typical segmentation algorithm, pixels are data to be processed by a central agent—the algorithm or the neural network trained as a pixel classifier. That LEGION is a massively parallel network of dynamical systems with mainly local coupling makes it particularly feasible for analog very large scale integration implementation, the success of which would be a major step toward real-time processing of scene segmentation.

A thorny issue with scene segmentation is that often no unique answer exists. A house, for example, may be grouped into a single segment if viewed afar. The same house, if viewed nearby, may be broken into multiple segments, including a door, a roof, windows, and other elements. This situation demands a flexible treatment of scene segmentation; a system should be able to generate multiple sets of segmentation easily, each of which should be reasonable. In LEGION, this flexibility is reflected to a certain degree by the effects of the parameters of \( W_z \) and \( \theta_p \), as discussed in section 5. As noted there, both Figure 9B and Figure 9C offer arguably reasonable results.

The LEGION network used so far has only one layer of oscillators. In section 4, we mentioned that the oscillatory dynamics of one-layer LEGION has a limited segmentation capacity (see Figure 6). It is interesting to note that the human perceptual system is also limited in simultaneously attending to the objects in a scene (Miller, 1956). We expect that the ability of LEGION
Image Segmentation improves significantly when multiple layers are used in subsequent stages. This multistage processing provides a natural way out of this fundamental limitation. In multistage processing, each layer does not need to segregate more than several segments, and yet the system as a whole can segregate many more segments than the segmentation capacity, an idea reminiscent of chunking proposed by Miller (1956). When the number of segments in an image is greater than the segmentation capacity, one-layer LEGION will produce a number of segments (simple or congregate) up to the segmentation capacity (see Figure 6). Congregate segments, however, can be further segmented with another layer of LEGION, whereas simple segments will not segment further. With multistage processing, the hierarchical system can provide results of both coarse- and fine-grain segmentation.

6.2 Biological Relevance. The relaxation-type oscillator used in LEGION is dynamically very similar to numerous other oscillators used in modeling neuronal behavior. Examples include the FitzHugh-Nagumo equations (FitzHugh, 1961; Nagumo, Arimoto, & Yoshizawa, 1962) and the Morris-Lecar model (Morris & Lecar, 1981). These can all be viewed as simplifications of the Hodgkin-Huxley equations (Hodgkin & Huxley, 1952). In equation 3.1, the variable \( x \) corresponds to the membrane potential of the neuron, and \( y \) corresponds to the channel activation or inactivation state variable, which evolves on the slowest time scale. The reduction from a full Hodgkin-Huxley model to the two-variable model is achieved by assuming that the other, faster, channel state variables are instantaneous. The dynamics of the lateral potential as given in equation 3.2 has properties similar to those of certain membrane channels and excitatory chemical synapses. The \( N \)-methyl-D-aspartate channel, for example, turns off on a slow time scale (Traub & Miles, 1991). Moreover, with a sufficiently large input, a cell with these channels can be transformed from the excitatory to the oscillatory mode. We note that the lateral potential does not act as a temporal integrator of all the input converging on its corresponding oscillator, but utilizes sharp nonlinearity as embodied by the outer Heaviside function in equation 3.2.

The theory of oscillatory correlation is consistent with the growing body of evidence that supports the existence of neural oscillations in the visual cortex and other brain regions. In the visual system, synchronous oscillations have been observed in cell recordings of the cat visual cortex (Eckhorn et al., 1988; Gray, König, Engel, & Singer, 1989). These neural oscillations are stimulus dependent and range from 30 to 70 Hz, often referred to as 40-Hz oscillations. Also, synchronous oscillations (locking with zero phase lag) occur across an extended brain region only if the stimulus constitutes a coherent object. These basic findings have been confirmed repeatedly in different brain regions and in different animal species (for reviews see Buzsáki, Llinás, Singer, Berthoz, & Christen, 1994; and Singer & Gray, 1995).

The local excitatory connections assumed in LEGION conform with various lateral connections in the brain. Relating to the visual cortex, these ex-
citatory connections, which link the excitatory elements of oscillators, could be interpreted as the horizontal connections in the visual cortex (Gilbert & Wiesel, 1989; Gilbert, 1992). It is known that horizontal connections originate from pyramidal cells, which are of excitatory type, and pyramidal cells are also the principal target of the horizontal connections. Furthermore, at the functional level, physiological recordings from monkeys suggest that motion-based visual segmentation may be processed in the primary visual cortex (Stoner & Albright, 1992; Lamme, van Dijk, & Spekreijse, 1993). The global inhibitor (see Figure 3) receives input from the entire oscillator network and feeds back inhibition onto the network. It serves to segment multiple patterns simultaneously present in a visual scene, thus exerting a global coordination. Crick (1984) has suggested that part of the thalamus, the thalamic reticular complex in particular, may be involved in the global control of selective attention. The thalamus is uniquely located in the brain; it receives input from and sends projections to almost the entire cortex. This suggestion and key anatomical and physiological properties of the thalamus prompt us to speculate that the global inhibitor might correspond to a neuronal group in the thalamus. The activity of the global inhibitor should be interpreted as the collective behavior of the neuronal group.

6.3 Figure-Ground Segregation. The dynamics proposed in this article separates a scene into a number of major segments and a background, which corresponds to the rest of the scene. The major segments combine to form the foreground, whose corresponding oscillators are oscillatory until the input scene fades away. After a brief beginning period, the oscillators corresponding to the background become excitable and stop oscillating. This dynamics effectively gets rid of noisy fragments without either prior smoothing or postprocessing of removing small regions, the methods often used in segmentation algorithms. With this dynamics, typical figure-ground segregation can be characterized as a special case, where only one major segment is allowed to be separated from the scene. In this sense, we claim that our dynamics also provides a potential solution to the problem of figure-ground segregation. We allow a foreground to include multiple segments so that both scene segmentation and figure-ground segregation are incorporated in a unified framework.

6.4 Future Topics. In this study, we have not addressed the role of prior knowledge in image segmentation. For example, when people segment the images of Figures 9A and 10A, they inevitably use their knowledge of human anatomy, which describes among other things the relative size and position of major brain regions. A more complete system of image segmentation must address this issue. Besides prior knowledge, many grouping principles outlined in section 1 have not been incorporated into the system. One of the main future topics is to incorporate more grouping cues into the system. The global inhibitory mechanism will play a key role in overall
system coordination; it makes various factors compete with each other, and
a final segment is formed because of strong binding within the segment.

Our study in this article focuses exclusively on visual segmentation. It
should be noted that neural oscillations occur in other modalities as well,
including audition (Galambos, Makeig, & Talmachoff, 1981; Ribary et al.,
1991) and olfaction (Freeman, 1978). Strikingly, these oscillations in dif-
ferent modalities show comparable frequencies. A recent study extended
LEGION to deal with auditory scene segregation (Wang, 1996). With its
computational properties and its biological relevance, the oscillatory corre-
lation approach promises to provide a general neurocomputational theory
for scene segmentation and perceptual organization.

Acknowledgments

We thank the two anonymous referees for their constructive comments.
D.L.W. was supported in part by ONR grant N00014-93-1-0335, NSF grant
IRI-9423312, and ONR YIP Award N00014-96-1-0676. D.T. was supported
in part by NSF grant DMS-9423796. We thank S. Campbell for preparing
Figure 2 and E. Cesmeli for his assistance in preparing Figure 7.

References

Cambridge University Press.
Barlow, H. B. (1972). Single units and cognition: A neurone doctrine for percept-
Eckhorn, R., Bauer, R., Jordan, W., Brosch, M., Kruse, W., Munk, M., & Reitboeck,
FitzHugh, R. (1961). Impulses and physiological states in models of nerve mem-
rectilinear segments by the labeled Hough transform. CVGIP: Image Under-
standing, 58(3), 22-42.
Freeman, W. J. (1978). Spatial properties of an EEG event in the olfactory bulb


Received April 1, 1996; accepted August 20, 1996.