Range Image Segmentation Using a Relaxation Oscillator Network

Xiuwen Liu, *Student Member, IEEE,* and DeLiang L. Wang, *Associate Member, IEEE*

Abstract— A locally excitatory globally inhibitory oscillator network (LEGION) is constructed and applied to range image segmentation, where each oscillator has excitatory lateral connections to the oscillators in its local neighborhood as well as a connection with a global inhibitor. A feature vector, consisting of depth, surface normal, and mean and Gaussian curvatures, is associated with each oscillator and is estimated from local windows at its corresponding pixel location. A context-sensitive method is applied in order to obtain more reliable and accurate estimations. The lateral connection between two oscillators is established based on a similarity measure of their feature vectors. The emergent behavior of the LEGION network gives rise to segmentation. Due to the flexible representation through phases, our method needs no assumption about the underlying structures in image data and no prior knowledge regarding the number of regions. More importantly, the network is guaranteed to converge rapidly under general conditions. These unique properties may lead to a real-time approach for range image segmentation in machine perception.

Index Terms—LEGION, oscillatory correlation, pulse-coupled neural networks, range image segmentation, relaxation oscillator.

I. INTRODUCTION

IMAGE SEGMENTATION has long been considered in machine vision as one of the fundamental tasks. Range image segmentation is especially important because depth is one of the most widely used cues in visual perception. Due to its practical importance, many techniques have been proposed for range image segmentation, and they can be roughly classified into four categories: 1) edge-based algorithms [3], [30], [44]; 2) region-based algorithms [1], [15], [19], [24], [40]; 3) classification-based approaches [2], [14], [15], [18], [22]; and 4) global optimization of a function [25].

Edge-based algorithms first identify the edge points that signify surface discontinuity using certain edge detectors, and then try to link the extracted edge points together to form surface boundaries. For example, Wani and Batchelor [44] introduced specialized edge masks for different types of discontinuity. Because critical points, such as junctions and corners, could be degraded greatly by edge detectors, they are extracted in an additional stage. Then surface boundaries are formed by growing from the critical points. As we can see, many application-specific heuristics must be incorporated

in order to design good edge detectors and overcome the ambiguities inherent in linking.

Region-based algorithms were essentially similar to region-growing and split-and-merge techniques for intensity images [46], but with more complicated criteria to incorporate surface normal and curvatures which are critical for range image segmentation. A commonly used method is iterative surface fitting [1], [15], [24]. Pixels are first coarsely classified based on the sign of mean and Gaussian surface curvature and seed regions are formed based on initial classification. Neighboring pixels will be merged into an existing surface region if they fit into the surface model well. This procedure is done iteratively. As pointed by Hoffman and Jain [14], the major disadvantage is that many parameters need to be involved. Also a good surface model that fits the range image data must be provided in order to obtain good results.

In classification-based approaches, range image segmentation is posed as a vector quantization problem. Each pixel is associated with an appropriate feature vector. The center vector for each class can be obtained by applying some clustering algorithms which minimize a certain error criterion [14] or alternatively by training [2], [22]. Then each pixel is associated with the closest cluster center. The segmentation result from the classification can be further refined by a merging procedure similar to region-growing [2], [14], [15], [22]. One of the limitations of classification-based approaches is that the number of regions must be given *a priori,* which, generally, is not available. Koh *et al.* [22] and Bhandarkar *et al.* [2] tried to address this issue by using a hierarchical self-organizing network for range image segmentation. At each level, a self-organizing feature map (SOFM) [23] is used to segment range images into a given number of regions. An *abstract tree* [22] is constructed to represent the output of the hierarchical SOFM network. The final segmentation is obtained by searching through the abstract tree, which is sequential and similar to a split-and-merge method. Thus the solution suffers from the disadvantages of region-based algorithms. In addition, the problem of prior specification of number of regions is not entirely solved because the number of regions for each level still needs to be specified.

A more fundamental limitation common to all region- and classification-based approaches is that the representation is too rigid, i.e., different regions are represented through explicit labels, which forces the approaches to be sequential to a large extent. Energy minimization techniques [9] can be inherently parallel and distributed and have been widely used in image classification and segmentation. In this framework, solutions are found by minimizing energy functions using relaxation

X. Liu is with the Department of Computer and Information Science, The Ohio State University, Columbus, OH 43210 USA.

D. L. Wang is with the Department of Computer and Information Science and the Center for Cognitive Science, The Ohio State University, Columbus, OH 43210 USA.

algorithms.[1] Li [25] constructed a set of energy functions for range image segmentation and recognition by incorporating surface discontinuity through mean and Gaussian curvatures. Minimization algorithms were obtained based on regularization techniques [38] and relaxation labeling algorithms [17], [34]. While the approach was quite successful, the main problem is that the algorithms are too computationally expensive for real-time applications [10], [28].

In this paper, we use a novel neural network for segmenting range images, which overcomes some of the above limitations. Locally excitatory globally inhibitory oscillator network (LEGION) [37], [42], [43] provides a biologically plausible framework for image segmentation in general. Each oscillator in the LEGION network connects excitatorily with the oscillators in its neighborhood as well as inhibitorily with a global inhibitor. For range image segmentation, the feature detector associated with each oscillator estimates the surface normal and curvatures at its corresponding pixel location. The lateral connection between two oscillators is set at the beginning based on a similarity measure between their feature vectors. The segmentation process is the emergent behavior of the oscillator network. Because the results are encoded flexibly through oscillator phases, segmentation is inherently parallel and no assumption about the underlying structures in image data is needed.

The rest of this paper is organized as follows. Section II gives an overview of the LEGION network. Section III presents feature vector estimation for range images. Section IV provides experimental results and comparisons with other approaches. Section V justifies the biological plausibility of this approach and gives a comparison with pulse-coupled neural networks (PCNN's). A preliminary version of this paper appeared in [27].

II. OVERVIEW OF THE LEGION DYNAMICS

A fundamental problem in image segmentation is to group similar input elements together while segregating different groups. Temporal correlation was hypothesized as a representational framework for brain functions [29], [41], which has received direct support from neurophysiological findings of stimulus-dependent oscillatory activities in the visual cortex [6], [12]. The LEGION network [37], [42], which is based on oscillatory correlation, was proposed as a computational model to address problems of image segmentation and figure-ground segregation [43]. In this section, after briefly introducing the single oscillator model, we summarize the properties of the network and demonstrate the network's capability using a synthetic image.

A. Single Oscillator Model

As the building block of LEGION, a single oscillator i is defined as a feedback loop between an excitatory unit u_i and

[1] "Relaxation" used in relaxation labeling technique [34] refers to an optimization technique that global optimal solutions can be obtained by satisfying local constraints. This is very different from the term "relaxation" as used in relaxation oscillators, where it refers to the change of activity on a slow time scale (see Section II).
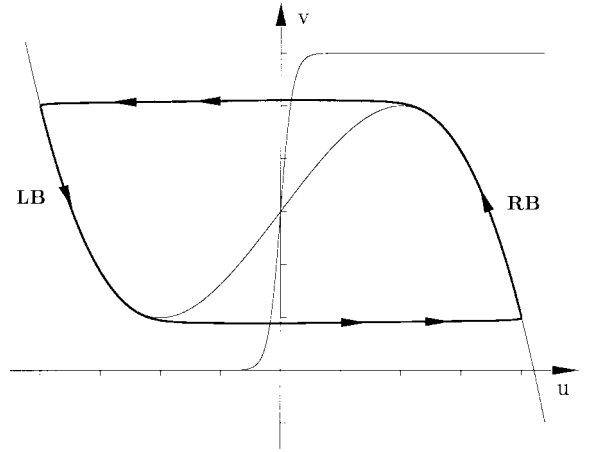

Fig. 1. A stable limit cycle for a single relaxation oscillator. The thick solid line represents the limit cycle and thin solid lines stand for nullclines. Arrows are used to indicate the different traveling speed, resulting from fast and slow time scales. The following parameter values are used: $\epsilon = 0.02$, $\beta = 0.1$, $\gamma = 3.0$, and a constant stimulus $I = 1.0$.

an inhibitory unit v_i [43]

$$\frac{du_i}{dt} = 3u_i - u_i^3 + 2 - v_i + I_i H(p_i - \theta) + S_i + \rho \quad (1a)$$

$$\frac{dv_i}{dt} = \epsilon(\gamma(1 + \tanh(u_i/\beta)) - v_i). \quad (1b)$$

Here H stands for the heaviside step function. I_i represents external stimulation for the oscillator i, which is assumed to be applied at time zero. ρ denotes a Gaussian noise term to test the robustness of the system and play an active role in desynchronization. S_i denotes the coupling from other oscillators in the network, and p_i the potential of the oscillator i. θ is a threshold, where $0 < \theta < 1$. The only difference between the definition in [43] and the one given here is that the term $H(p_i + \exp(-\alpha t) - \theta)$ in [43] is replaced by $H(p_i - \theta)$ in (1a) so that the Heaviside term in (1) does not depend explicitly on time.

The parameter ϵ is a small positive number. In this case, (1), without any coupling or noise but with constant stimulation, corresponds to a standard relaxation oscillator, similar to the van der Pol oscillator [39]. The u-nullcline ($du/dt = 0$) of (1) is a cubic curve, and the v-nullcline ($dv/dt = 0$) is a sigmoid function, the steepness of which is specified by β. If $I > 0$ and $H = 1$, which corresponds to a positive stimulation, (1) produces a stable periodic orbit for all sufficiently small values of ϵ, alternating between left and right branches of near steady-state behavior. Fig. 1 shows a stable limit cycle along with nullclines. The oscillator travels on left and right branches on a slow time scale because the driving force is mainly from the inhibitory unit and is weak, while the transition between two branches occurs on a fast time scale because the driving force is mainly from the excitatory unit and is strong. The slow and fast time scales result from a small ϵ and produce highly nonlinear activities.

The parameter γ is used to control the ratio of the times that an oscillator spends on the right and left branches. Fig. 2 shows the temporal activities of the excitatory unit for $\gamma = 3$ and $\gamma = 40$. If γ is chosen to be large, the output will be close to neural spikes. Based on this observation, the relaxation oscillator can also be viewed as a neuronal spike generator.
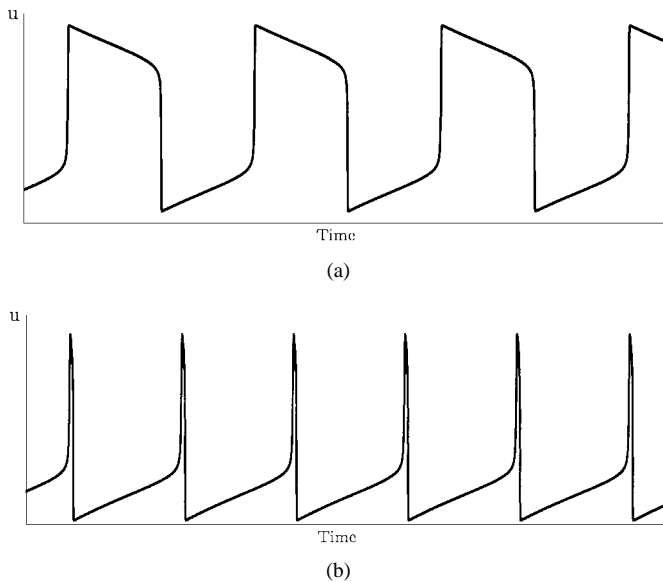
Fig. 2. The temporal activities of the excitatory unit of a single oscillator for different γ values. Other parameters are same as for Fig. 1: (a) $\gamma = 3.0$ and (b) $\gamma = 40.0$.

B. Emergent Behavior of LEGION Networks

The LEGION network consists of identical relaxation oscillators given by (1) and a global inhibitor. Each oscillator is excitatorily coupled with oscillators in its local neighborhood. The coupling term S_i is defined as

$$S_i = \sum_{k \in N_c(i)} W_{ik} H(u_k - \theta_u) - W_z H(Z - \theta_z). \quad (2)$$

Here θ_u and θ_z are thresholds. W_{ik} is the connection weight from oscillator k to i, and $N_c(i)$, the coupling neighborhood of i, is a set of neighboring oscillators of i. An oscillator only sends excitation to its neighbors when its activity is higher than the threshold θ_u. In that case, we refer that it is in *active* phase. Otherwise, it is in *silent* phase. This results in heaviside coupling.

With excitatory coupling, the distance between two oscillators decreases at an exponential rate when traveling on the same branch. During jumps between left and right branches, the *time difference,* the time that is needed to travel the distance between them, will be compressed even though the Euclidean distance does not change. These two factors lead to the fact that the two oscillators get synchronized at an exponential rate [37], [42], [43].

The global inhibitor Z leads to desynchronization among different oscillator groups [37], [42]. In (2), W_z is the weight of inhibition from the global inhibitor Z, which is defined by

$$\frac{dZ}{dt} = \phi(\sigma_\infty - Z) \quad (3)$$

where $\sigma_\infty = 1$ if $u_i \geq \theta_z$ for at least one oscillator i, and $\sigma_\infty = 0$ otherwise. Under this condition, only an oscillator in the active phase can trigger the global inhibitor. When one oscillator group is active, it suppresses the other groups from jumping up by activating the global inhibition. Architecturally, the global inhibitor imposes constraints on the entire network
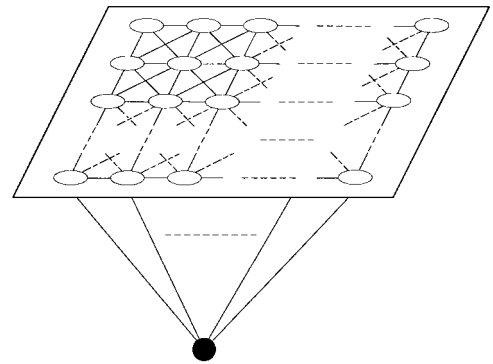

Fig. 3. Architecture of a two-dimensional LEGION network with eight-nearest neighbor coupling. An oscillator is indicated by an empty ellipse, and the global inhibitor is indicated by a filled circle.

and effectively reduces the collisions from local coupling. This is a main reason why LEGION networks are far more efficient than purely locally coupled cooperative processes, including relaxation labeling techniques [9], [17], [34], which tend to be very slow [10], [28].

The network we use for image segmentation is two dimensional. Fig. 3 shows the architecture employed throughout this paper, where each oscillator is connected with its eight immediate neighbors, except those on the boundaries where no wraparound is used. To allow the network to distinguish between major regions and noisy fragments, the *potential* of the oscillator i, p_i, is introduced. The basic idea is that a major region must contain at least one *leader*. A leader is an oscillator that receives large enough lateral excitation from its neighborhood, i.e., larger than θ_p (called the *potential threshold*). Such a leader can develop a high potential and lead the activation of the oscillator block corresponding to an object. A noisy fragment does not contain such an oscillator and thus is not able to be active after a beginning period. See [43] for detailed discussion.

We illustrate the capability of the network for image segmentation using a binary synthetic image. The original image is shown in Fig. 4(a). It has 40×40 pixels and consists of seven geometric objects. Fig. 4(b) shows the corrupted version by adding 10% noise: each pixel has 10% chance of flip-flop. The corrupted image is presented to a 40×40 LEGION network. Each oscillator in the network starts with a random phase, shown in Fig. 4(c). Fig. 4(d)–(j) shows the snapshots of the network activity when the oscillators corresponding to one object are in the active phase. All the oscillator groups corresponding to the objects are popped out alternately after five cycles. Fig. 5 shows the temporal activities of all the stimulated oscillators, grouped together by objects, and a background, which consists of those stimulated oscillators that do not correspond to any object. From Fig. 5, we can see that desynchronization among seven groups as well as synchronization within each object is achieved quickly, here after five cycles. Furthermore, only the oscillators belonging to an object can stay oscillating and all other noisy fragments are suppressed into the background after two cycles.

Formally, it has been proved [37], [43] that the LEGION network can achieve synchronization among the oscillators
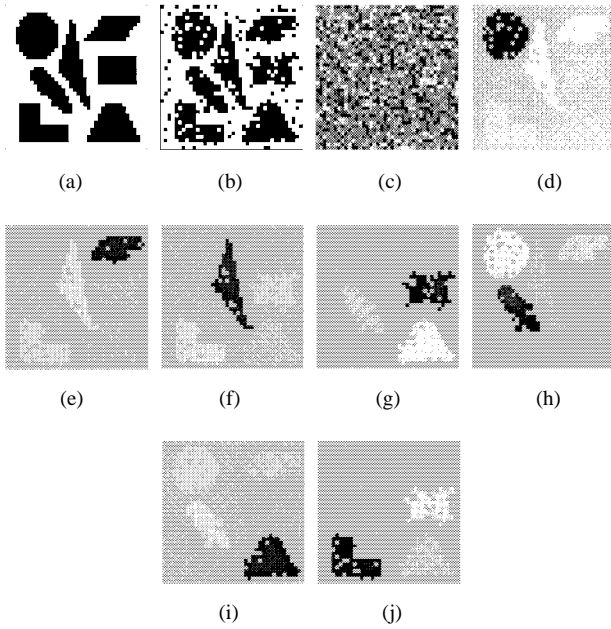
Fig. 4. Illustration of LEGION dynamics: (a) an input image consisting of seven geometric objects, with 40 × 40 pixels, (b) the corrupted image of (a) by adding 10% noise, which is presented to a 40 × 40 LEGION network, (c) a snapshot of the network activity at the beginning, and (d)–(j) Subsequent snapshots of the network activity. In (c)–(j), the grayness of a pixel is proportional to the corresponding oscillator's activity and black pixels represent oscillators in the active phase. The parameter values for this simulation are following: $\epsilon = 0.02$, $\beta = 0.1$, $\gamma = 20.0$, $\theta_u = -0.5$, $\theta_p = 7.0$, $\theta_z = 0.1$, $\theta = 0.8$, and $W_z = 2.0$.

corresponding to the same object and desynchronization between groups corresponding to different objects quickly under general conditions. In particular, synchronization within an object and desynchronization among different objects can be achieved in $N_M + 1$ cycles, where N_M is the number of the major regions corresponding to distinct objects.

III. SIMILARITY MEASURE FOR RANGE IMAGES

Given the LEGION dynamics, the main task is to establish lateral connections based on some similarity measure. Intuitively, neighboring pixels belonging to the same surface should be similar and thus should have strong connections with each other in the LEGION network. To be effective for range image segmentation, we consider different types of surfaces. For planar surfaces, surface normal is generally homogeneous while depth value may vary largely. For cylindrical, conical, and spherical surfaces, surface curvature does not change much while both depth value and surface normal may. Based on these observations, the similarity measure should depend on surface normal and curvatures in addition to raw depth data. Similar to [22], we chose (z, n_x, n_y, n_z, C, G) as the feature vector for each oscillator. Here z is the depth value, (n_x, n_y, n_z) is the surface normal, and C and G are the mean and Gaussian curvature, at the corresponding pixel location. Depth value is directly available from image data. Surface normal at each pixel is estimated from the depth values of a neighboring window, and curvatures are further derived from surface normal.

Formally, each oscillator is associated with a feature detector to estimate the normal and curvature values at the corresponding pixel location. Based on the work in [1], (n_x, n_y, n_z) is calculated from the first-order partial derivatives

$$(n_x, n_y, n_z) = \frac{\frac{\partial z}{\partial x} \times \frac{\partial z}{\partial y}}{\left\| \frac{\partial z}{\partial x} \times \frac{\partial z}{\partial y} \right\|} \tag{4}$$

and the partial derivatives are estimated using the following formulas[2]:

$$\frac{\partial z}{\partial x} \approx D_x = d_0 \cdot d_1^T \tag{5a}$$

$$\frac{\partial z}{\partial y} \approx D_y = d_1 \cdot d_0^T. \tag{5b}$$

Here $\cdot$ is the inner product operation and T indicates transpose, d_0 is the equally weighted average operator, and d_1 is the least-square derivative estimator. For a 5×5 window, d_0 and d_1 are given by

$$d_0 = \tfrac{1}{5}[1\ 1\ 1\ 1\ 1]^T \tag{6a}$$

$$d_1 = \tfrac{1}{10}[-2\ -1\ 0\ 1\ 2]^T. \tag{6b}$$

This normal estimation method works well if the estimation window is within one surface. When the window crosses different surfaces, especially ones with very different depth values, the estimation results tend to be inaccurate. In order to improve the results near surface boundaries, we require that the pixels in the estimation window must be within the same *context* based on the edge preserving noise smoothing quadrant filter [16] and the context-sensitive normal estimation method [45]. However, both methods require edge information, which is not available for range images. To be more applicable, here we define that two pixels are within the same context if their difference in depth value is smaller than a given threshold. This definition captures the significant edge information in range images and works well. When estimating the first-order derivatives, d_0 and d_1 are applied only to pixels that are within the same context with respect to the central pixel. These operators are called *context-sensitive* operators.

There are two ways to estimate the surface curvature. First it can be estimated by breaking the surface up into curves and measuring the curvature of these curves. Second it can be estimated by measuring surface normal changes. Following [18], the surface curvature at point $\mathbf{p}$ in the direction of point $\mathbf{q}$ is estimated by

$$\kappa_{p,q} = \begin{cases} \dfrac{\|\mathbf{n}_p - \mathbf{n}_q\|}{\|\mathbf{p} - \mathbf{q}\|}, & \text{if } \|\mathbf{p} - \mathbf{q}\| \leq \|(\mathbf{n}_p + \mathbf{p}) \\ & \quad -(\mathbf{n}_q + \mathbf{q})\| \\ -\dfrac{\|\mathbf{n}_p - \mathbf{n}_q\|}{\|\mathbf{p} - \mathbf{q}\|}, & \text{otherwise} \end{cases} \tag{7}$$

where $\mathbf{p}$ and $\mathbf{q}$ refer to the 3-D coordinate vectors of the corresponding pixels, and $\mathbf{n}_p$ and $\mathbf{n}_q$ are the unit normal vectors at points $\mathbf{p}$ and $\mathbf{q}$, respectively, which are available from (4). Here the 3-D coordinate vector of a pixel is composed of its

---

[2]Here the formulas are implemented using inner product operations. For example, in (5a), inner products between d_1^T and the rows in the local window are taken first, and the estimation result is the inner product between d_0 and the resulting column vector.
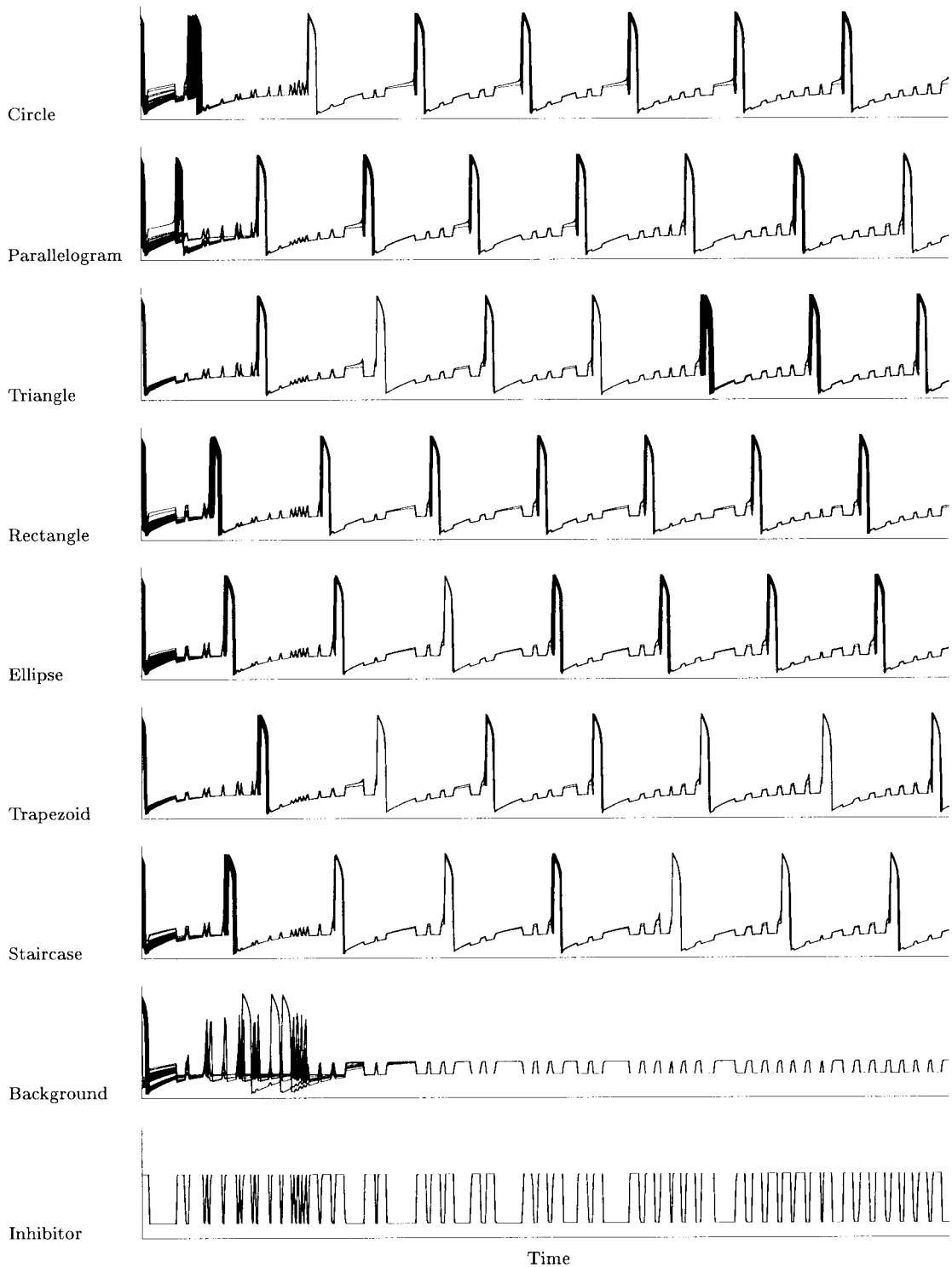
Fig. 5.   Temporal evolution of the LEGION network. The upper seven plots show the combined temporal activities of the seven oscillator blocks representing the corresponding geometric objects. The eighth plot shows the temporal activities of all the stimulated oscillators which correspond to the background. The bottom one shows the temporal activity of the global inhibitor. The simulation took 20 000 integration steps using a fourth-order Runge–Kutta method to solve differential equations.

2-D location in the image and its depth value. The condition in (7) is to assign a positive curvature value for pixels on a convex surface and a negative one for pixels on a concave surface.

Now the mean curvature C_i of oscillator i, is estimated as the average of all possible values with respect to a neigh-

borhood $N_k(i)$ of oscillator i. The Gaussian curvature G_i is estimated as the production of the minimum and maximum values with respect to the neighborhood.

In summary, the similarity measure between two oscillators is given by: $W_{ik} = 255/(\Psi(i, k) + 1)$, and the lateral connections are set up accordingly. Here $\Psi(i, k)$ is a disparity

measure between oscillators i and k, and is given by

$$\Psi(i,\,k) = \begin{cases} K_z|z_i - z_k| + K_n\|\mathbf{n}_i - \mathbf{n}_k\| & \text{if } |C_i| \leq T_c \\ K_c|C_i - C_k| + K_G|G_i - G_k| & \text{otherwise.} \end{cases} \tag{8}$$

Here K_z, K_n, K_c, and K_G are weights for different disparity measures. T_c is a threshold for planar surface testing.

IV. EXPERIMENTAL RESULTS

For real images with a large number of pixels, it involves very intensive numerical computation to solve the differential equations of (1)–(3) if the LEGION network is applied directly. To reduce the numerical computation on a serial computer, an algorithm is extracted from these equations [43]. The algorithm follows the major steps of the original network and exhibits the essential properties of relaxation oscillator networks. It has been applied to intensity image segmentation [43].

A. Parameter Selection

Most parameters in (1)–(3) are intrinsic to LEGION networks and need not to be changed once they are appropriately chosen, and only the global inhibition W_z and the potential threshold θ_p need to be tuned for applications. Other parameters are application-specific and related to how to measure the similarity between feature vectors. Theoretically, best parameter values could be obtained through training using a neural network. Experimentally, each parameter roughly corresponds to a threshold for a type of discontinuity, and can be set accordingly. In (8), K_z captures abrupt changes in depth, and K_n, K_c, and K_G capture boundaries between different type surfaces.

There are several local windows involved. The system is not sensitive to the size of windows and can be chosen from a wide range. For the experiments in this work, we use a single system configuration. That is, the surface normal is estimated from a 5×5 neighboring window, the curvatures are estimated from a 3×3 neighboring window, and each oscillator has lateral connections with its eight immediate neighbors as depicted in Fig. 3.

B. Results

From an algorithmic point of view, a system for image segmentation must incorporate discontinuity in input data which gives rise to region boundaries. For range images, there are mainly three types of discontinuity. Jump edges occur when depth values are discontinuous due to occlusion while crease and curvature edges occur when different surfaces intersect with each other. Crease edges correspond to surface normal discontinuity while curvature edges are due to curvature discontinuity where surface normals are continuous. Lateral connection weights implicitly encode the discontinuity through (8), which will be demonstrated using real range images.

Range images shown in Figs. 6–10 were obtained from the Pattern Recognition and Image Processing Laboratory at Michigan State University. These images were generated using a Technical Arts 100X white scanner, which is a triangulation-based laser range sensor. Except for the global inhibition W_z
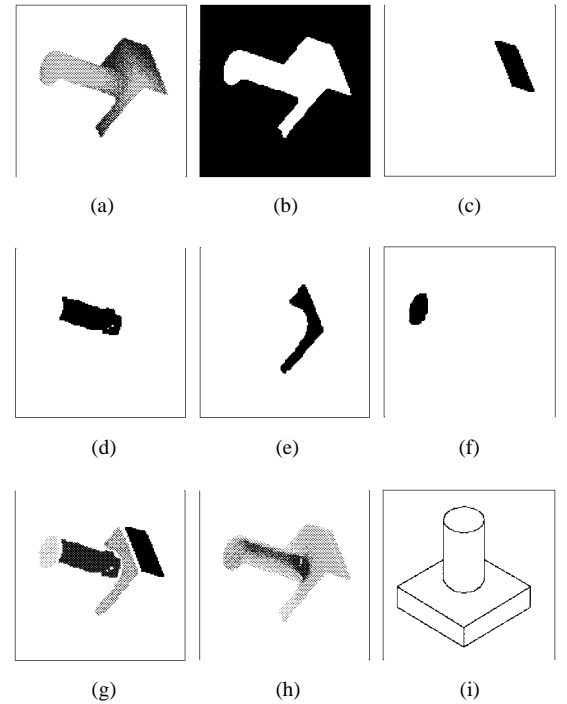

Fig. 6. Segmentation result of the LEGION network for the range image of a column: (a) the input range image, (b) the background region, (c)–(f) the four segmented regions, (g) the overall segmentation result represented by a gray map, (h) the corresponding intensity image, and (i) the 3-D construction model. As in Fig. 4, black pixels in (b)–(f) represent oscillators that are in the active phase.

and the potential threshold θ_p, which define a scale space (see below), all other parameters are in general fixed, as is the case for all the range images shown in this paper. For the range images shown here, a fixed level of global inhibition and potential threshold works well except for images with thin regions like those shown in Fig. 9. Because the image size is rather small compared to a 5×5 normal estimation window, a popped-out region is extended by one pixel within the same context. This simple postprocessing is done only for reducing the boundary effect in the final results.

Fig. 6 shows a complete example from our network. Fig. 6(a) shows a range image of a column. From its 3-D object model (the original 3-D construction) shown in Fig. 6(i), it consists of a cylinder and a rectangular parallelepiped. From the view point where the image is taken, four planar surfaces and a cylindrical one are visible. One planar surface is missing from the range image due to sampling artifact and shadows. Fig. 6(b)–(f) shows the output of the LEGION network. We can see that all the visible surface regions in the image, including the background, are popped up. Oscillators belonging to the same surface region are synchronized due to the strong lateral connections resulting from similar feature vectors and thus are grouped together correctly. The segregation of these regions is achieved because of the weak lateral connections resulting from jump and crease edges. Fig. 6(g) shows the overall result by putting the popped up regions together into a *gray map* [43], where each region is shown using a single gray level. The boundaries between different surfaces are shown fairly accurately, demonstrating that surface discontinuity is effectively captured through lateral
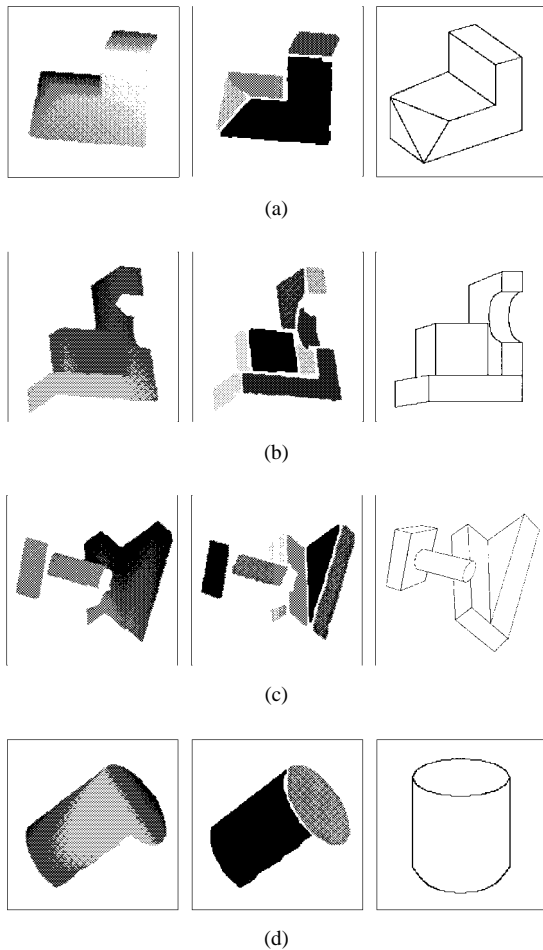
(a)

(b)

(c)

(d)

Fig. 7. Segmentation results of the LEGION network for range images. In each row, the left frame shows the input range image, the middle one shows the segmentation result represented by a gray map, and the right one shows the 3-D construction model for comparison purposes.

(a)

(b)

(c)

Fig. 8. Segmentation results of the LEGION network for several more range images. See the caption of Fig. 7 for arrangement.

(a)

(b)

Fig. 9. Two examples with thin regions. The global inhibition and potential threshold are tuned to get the results shown here. See the caption of Fig. 7 for arrangement.

connection. Small holes in Fig. 6(d) are due to some noise in the input range image. Here Fig. 6(h), the intensity image, is included for discussion in Section V.

Figs. 7 and 8 show the segmentation results for a number of real range images, which include different types of edges and junctions. These results were produced using the same parameter values as in Fig. 6. In Fig. 7(a), an object with only planar surfaces is segmented into four surface regions. Boundaries between surfaces are precisely marked, showing that crease edges between planar surfaces are correctly represented. The junction point where three surfaces intersect is handled correctly without additional modeling, which would be difficult for edge-based approaches. Fig. 7(b) shows an object with several planar surfaces and a cylindrical one. As in Fig. 7(a), all the planar surfaces are segmented precisely. The cylindrical surface is segmented correctly even though it is not complete because of shadowing effect. Both jump and crease edges are marked correctly in the result. Fig. 7(c) shows an image which consists of two objects. All the surfaces are correctly segmented. Fig. 7(d) shows a cylinder. The transition between two surfaces is smooth and it is even difficult to segment it manually. The system correctly segmented two surface regions and the boundary between them is marked where we would expect. Fig. 8(a) shows another image with smooth transition. Because of that, curvatures were used to segregate the surface
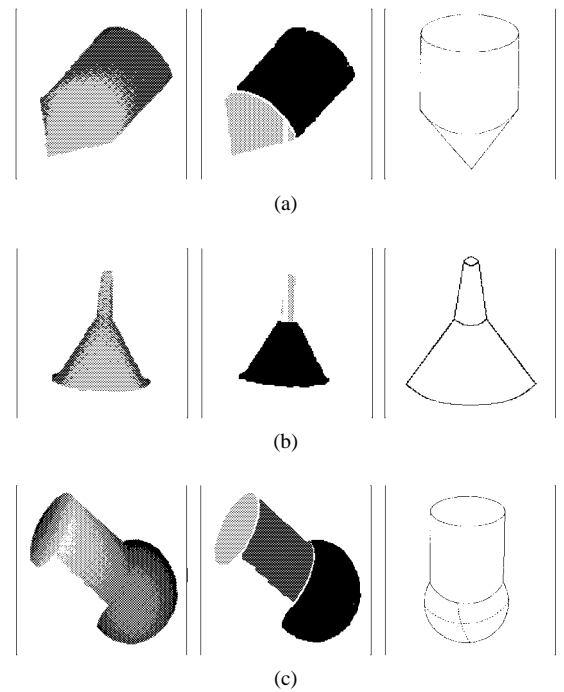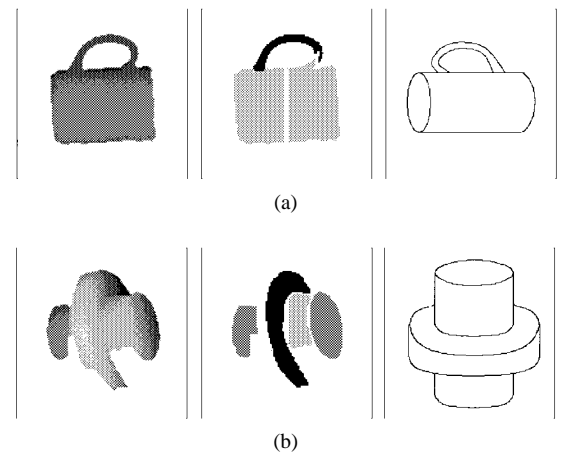
regions. Fig. 8(b) shows a funnel. Even though the neck is a thin surface, it is segmented correctly. But boundary effect is more obvious. Fig. 8(a) and (b) demonstrated that the system can handle conic surfaces. Fig. 8(c) showed that the network can handle spherical surfaces.

Fig. 9 shows two difficult examples with thin regions compared to the size of the normal estimation windows. When using the same parameters as in Figs. 7 and 8, the results are not very satisfactory. This is because the thin parts do not contain a leader and the normal and curvature estimations are not very reliable due to the smooth transitions between surfaces. When tuning the potential threshold and global inhibition, both images are processed correctly. All the surfaces are segmented with boundaries marked correctly, as shown in Fig. 9.

These examples show clearly that the LEGION network can handle planar, cylindrical, conic, and spherical surfaces and
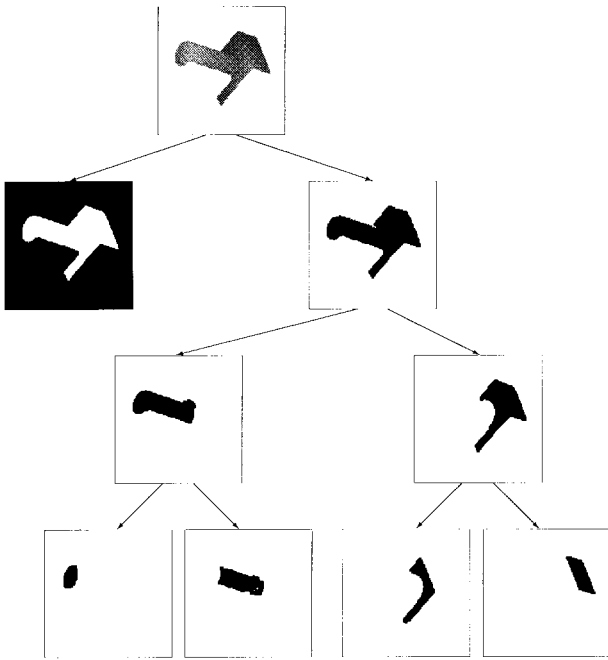
Fig. 10. A hierarchy obtained from multiscale segmentation. The top is the input range image and each segmented region is further segmented by increasing the level of global inhibition. As in Fig. 6, black pixels represent active oscillators, corresponding to the popped up region. See Fig. 6(i) for the corresponding 3-D model.

different types of surface discontinuity. Only a few parameters need to be tuned for all these images, which demonstrates that our network is a robust approach to range image segmentation.

C. Comparison with Existing Approaches

As demonstrated using real range images, lateral connections in our LEGION network capture different types of discontinuity as well as similarity within each surface region. Both factors are critically important for correct segmentation. Discontinuity avoids under-segmentation while similarity avoids over-segmentation. Critical points, such as junctions between surfaces are handled correctly due to the context-sensitive estimation. These factors determine that our method is close to edge-based approaches [3], [30], [44]. For region- and classification-based approaches [1], [2], [14], [22], [24], because in most cases many unspecified parameters and pre- and postprocessing are involved, a quantitative comparison suggested by Hoover *et al.* [15] is not possible. A qualitative comparison for similar images is used instead, which may be suggestive. For the range images used in this paper, we believe that our results in general are comparable with the best results from other methods. For the two images whose similar versions are also used elsewhere, we give a more detailed comparison below.

Cup images similar to Fig. 9(a) were also used in [1], [14], and [24]. In [14], a cup image was first classified into two patch images through three different tests. The two patch images were merged to form another patch image, based on which the final classification was generated. In the final result, the handle was classified as background, resulting in wrong topology. In [1], a coffee mug was segmented into six bivariate polynomial surfaces with complicated coefficients, which were obtained by

surface fitting after curvature-based initial classification. The handle was incorrectly segmented into two surfaces. In [24], a slightly different cup image was segmented into five parts through an iterative regression method, where the handle was broken into three parts. In our case, Fig. 9(a) shows that the cup was segmented correctly into two regions, the body and the handle. This qualitative comparison suggests that only our approach produced a correct segmentation, without under- or over-segmentation.

An image similar to Fig. 6(a) was also used in [2]. At the lower layers of the hierarchical network, the image was broken into small patches because a position term, not relevant to surface properties, is included in the feature vector in order to achieve spatial connectivity [2], [22]. Due to the position term, even planar surfaces need several layers to be correctly segmented. More complicated surfaces, such as cylindrical and conic ones, tend to be broken into small regions. As shown in [2], an image may be correctly classified only when the correct number of regions in a layer is given. The problem is not solved even using a hierarchical network. To produce meaningful results, segmentation from different layers must be combined. Because of that, the neural network based approaches [2], [22] in general became a split-and-merge algorithm.

Methodologically, our approach has several major advantages over other methods. First, the segmentation result is the emergent behavior of the LEGION network; the flexible representation of temporal labeling promises to provide a generic solution to image segmentation. Different cues, such as depth and color [see Fig. 6(a) and (h)] are highly correlated and should be integrated to improve the performance of a machine vision system. LEGION provides a unified and effective framework to implement the integration through lateral connections. On the other hand, the integration would be difficult for segmentation algorithms [1], [2], [14], [18], [22], [44]; different information must be modeled explicitly and encoded accordingly. Second, due to the nonlinearity of single oscillators and the effective architecture for modeling local and global coupling, our network is guaranteed to converge very efficiently. To our best knowledge, this is the only oscillatory network that has been analyzed rigorously. Finally, our approach is a dynamic system with parallel and distributed computation. The unique property of neural plausibility makes it particularly feasible for analog VLSI implementation, the efficiency of which may pave the way to achieve real-time range image segmentation.

V. DISCUSSIONS

A LEGION network has been constructed and applied to range image segmentation. Our network successfully segments real range images, which shows that it may give rise to a generic range image segmentation method.

A. Biological Plausibility of the Network

The relaxation oscillator used in LEGION is similar to many oscillators used in modeling neuronal behaviors, including FitzHugh–Nagumo [8], [32] and Morris–Lecar model [31], and is reducible from the Hodgkin–Huxley equations [13]. The local excitatory connections are consistent with various

lateral connections in the brain and can be viewed as the horizontal connections in the visual cortex [11]. The global inhibitor, which receives input from the entire network and feeds back with inhibition, can be viewed as a network that exerts some form of global control. Oscillatory correlation, as a special form of temporal correlation [29] [41], also conforms with experiments of stimulus-dependent oscillations, where synchronous oscillations occur when the stimulus is a coherent object, and no phase locking exists when regions are stimulated by different stimuli [6], [12], [35].

As stated earlier, the LEGION network can produce segmentation results within a few cycles. This analytical result may suggest a striking agreement with human performance in perceptual organization. It is well-known that human subjects can perform segmentation (figure-ground segregation) tasks in a small fraction of a second [4], corresponding to several cycles of oscillations if the frequency is taken to be around 40 Hz as commonly reported. Thus, if our oscillator network is implemented in real-time hardware, the time it takes the network to perform a segmentation task would be compatible with human performance.

B. Comparison with Pulse-Coupled Neural Networks

Both LEGION and PCNN networks were proposed based on the temporal correlation theory [29], [41] and experimental data of stimulus-dependent oscillations [6], [12]. Single units in both models approximate nonlinear neuronal activity, and excitatory local coupling is used to achieve synchronization. These are similarities between them. Yet there are important differences.

A single neuron in PCNN networks consists of input units, a linking modulation, and a pulse generator, all of which are mainly implemented using leaky integrators [7]. The input units consist of two principal branches, called *feeding* and *linking* inputs. Feeding inputs are the primary inputs from the neuron's receptive field while linking inputs are mainly the lateral connections with neighboring neurons and may also include feedback controls from higher layers. The linking modulation modulates the feeding inputs with linking inputs and produces the total internal activity U, which is given by $U = F(1 + bL)$. Here F and L are total feeding and linking inputs, respectively, and b is a parameter that controls the linking strength. The pulse generator is a Heaviside function of the difference between the total internal activity and a dynamic threshold implemented through another leaky integrator. If b is set to zero, the neuron becomes an encoder, which maps input values linearly to firing frequencies. If the activities from all neurons are summed up, the resulting output is a time series, which corresponds to the histogram of the input image [20]. With linking inputs, the behavior of the network can be very complicated. But when b is small, corresponding to the weak linking regime [20], synchronization is achieved only when the differences are small, similar to a smoothing operation. This type of PCNN can transform the input image into firing frequency representation with desirable invariant properties [20]. Variations of this type of PCNN have found successful applications in image factorization [21] and may have a connection with wavelet and other transformations [33].

Recently, Stoecker *et al.* [36] also applied PCNN to scene segmentation. The network successfully separated identical objects in a visual scene. Because different intensities map to different firing frequencies, the network would have difficulty in handling real images. For real images, different objects generally correspond to different frequencies, and the readout would be problematic, i.e., it would be difficult to have a network that could detect the results of segmentation. More fundamentally, the interactions among different firing frequencies have not been investigated. In [36], desynchronization is achieved solely due to the spatial distances among input objects. On the other hand, LEGION networks can successfully segment real range images as demonstrated here and other real images such as texture images [5].

Another difference is that LEGION has a well-analyzed mathematical foundation [37], [43] whereas there is little analytical investigation on PCNN, probably because the neuron model is complicated. As illustrated in this paper, to achieve image segmentation, complicated oscillator models may not be necessary. The task performed in [36] was previously achieved readily by LEGION [37], [42].

C. Further Research Topics

While the LEGION network successfully segmented real range images, there are several issues that need to be addressed in the future. One direct problem is how to train the LEGION network so that the optimal parameter values can be chosen automatically. One solution would be to have a separate network. A better solution should utilize the temporal dynamics and devise an efficient memory accordingly. A substantial improvement of the segmentation results is possible by setting the lateral connections according to the temporal context that is developed through dynamics. This offers another dimension which is unique to dynamic system approaches.

It is obvious that optimal segmentation is scale-dependent, and global inhibition in LEGION actually defines a scale space. Fig. 10 shows a hierarchy by manually changing global inhibition. At the first level, it corresponds to figure-ground segregation. By increasing global inhibition, the segmented results are refined further. As shown Fig. 10, this scale space is not based on blurring effects by continuously changing the filter's spatial extent [26]. Rather, the boundaries are precisely located through all levels. This scale space may be an alternative to conventional scale spaces and its properties need to be further studied.

REFERENCES

[1] P. J. Besl and R. C. Jain, "Segmentation through variable-order surface fitting," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. 10, pp. 167–192, 1988.
[2] S. M. Bhandarkar, J. Koh, and M. Suk, "Multiscale image segmentation using a hierarchical self-organizing map," *Neurocomput.,* vol. 14, pp. 241–272, 1997.
[3] B. Bhanu, S. Lee, C. C. Ho, and T. Henderson, "Range data processing: Representation of surfaces by edges," in *Proc. IEEE Int. Pattern*

Recognition Conf., 1986, pp. 236–238.

[4] I. Biederman, "Recognition-by-component: A theory of human image understanding," *Psych. Rev.,* vol. 94, pp. 115–147, 1987.

[5] E. Cesmeli and D. L. Wang, "Texture segmentation using Gaussian Markov random fields and LEGION," in *Proc. IEEE Int. Conf. Neural Networks,* 1997, vol. 3, pp. 1529–1534.

[6] R. Eckhorn, R. Bauer, W. Jordan, M. Brosch, W. Kruse, M. Munk, and H. J. Reitboeck, "Coherent oscillations: A mechanism of feature linking in the visual cortex?," *Biol. Cybern.,* vol. 60, pp. 121–130, 1988.

[7] 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.,* vol. 2, pp. 293–307, 1990.

[8] R. FitzHugh, "Impulses and physiological states in models of nerve membrane," *Biophys. J.,* vol. 1, pp. 445–466, 1961.

[9] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. 6, pp. 721–741, 1984.

[10] Z. Gigus and J. Malik, "Detecting curvilinear structure in images," Comput. Sci. Division, Univ. California, Berkeley, Tech. Rep. UCB/CSD 91/619, 1991.

[11] C. D. Gillbert, "Horizontal integration and cortical dynamics," *Neuron,* vol. 9, pp. 1–13, 1992.

[12] C. M. Gray, P. König, A. K. Engel, and W. Singer, "Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties," *Nature,* vol. 338, pp. 334–337, 1989.

[13] A. L. Hodgkin and A. F. Huxley, "A quantitative description of membrane current and its application to conduction and excitation in nerve," *J. Phys.,* vol. 117, pp. 500–544, 1952.

[14] R. Hoffman and A. K. Jain, "Segmentation and classification of range images," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. 9, pp. 608–620, 1987.

[15] A. Hoover and G. Jean-Baptiste, *et al.,* "An experimental comparison of range image segmentation algorithms," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. 18, pp. 673–689, 1996.

[16] J. Y. Hsiao and A. A. Sawchuk, "Unsupervised textured image segmentation using feature smoothing and probabilistic relaxation techniques," *Comput. Vision, Graphics, Image Processing,* vol. 48, no. 1, pp. 1–21, 1989.

[17] R. H. Hummel and S. W. Zucker, "On the foundations of relaxation labeling processes," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. 5, pp. 267–286, 1983.

[18] D. J. Ittner and A. K. Jain, "3-D surface discrimination from local curvature measures," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognition,* 1985, pp. 119–123.

[19] X. Y. Jiang and H. Bunke, "Fast segmentation of range images into planar regions by scan line grouping," *Machine Vision Applicat.,* vol. 7, no. 2, pp. 115–122, 1994.

[20] J. L. Johnson, "Pulse-coupled neural nets: Translation, rotation, scale, distortion, and intensity signal invariance for images," *Appl. Opt.,* vol. 33, no. 26, pp. 6239–6253, 1994.

[21] J. L. Johnson, M. L. Padgett, and W. A. Friday, "Multiscale image factorization," in *Proc. IEEE Int. Conf. Neural Networks,* 1997, vol. 3, pp. 1465–1468.

[22] J. Koh, M. Suk, and S. M. Bhandarkar, "A multilayer self-organizing feature map for range image segmentation," *Neural Networks,* vol. 8, no. 1, pp. 67–86, 1995.

[23] T. Kohonen, *Self-Organizing Maps.* Berlin, Germany: Springer-Verlag, 1995.

[24] A. Leonardis, A. Gupta, and R. Bajcsy, "Segmentation as the search for the best description of the image in terms of primitives," in *Proc. Int. Conf. Comput. Vision,* 1990, pp. 121–125.

[25] S. Z. Li, "Toward 3D vision from range images: An optimization framework and parallel networks," *Comput. Vision, Graphics, Image Processing: Image Understanding,* vol. 55, no. 3, pp. 231–260, 1992.

[26] T. Linderberg, *Scale-Space Theory in Computer Vision.* Dordrecht, the Netherlands: Kluwer, 1994.

[27] X. Liu and D. L. Wang, "Range image segmentation using an oscillatory network," in *Proc. IEEE Int. Conf. Neural Networks,* 1997, vol. 3, pp. 1656–1660.

[28] D. Marr, *Vision.* New York: Freeman, 1982.

[29] P. M. Milner, "A model for visual shape recognition," *Psych. Rev.,* vol. 81, no. 6, pp. 521–535, 1974.

[30] A. Mitiche and J. K. Aggarwal, "Detection of edges using range information," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. 5, pp. 174–178, 1983.

[31] C. Morris and H. Lecar, "Voltage oscillations in the barnacle giant muscle fiber," *Biophys. J.,* vol. 35, pp. 193–213, 1981.

[32] J. Nagumo, S. Arimoto, and S. Yoshizawa, "An active pulse transmission

line simulating nerve axon," in *Proc. Inst. Radio Eng.,* pp. 2061–2070, vol. 50, 1962.

[33] M. L. Padgett and J. L. Johnson, "Pulse coupled neural networks (PCNN) and wavelets: Biosensor applications," in *Proc. IEEE Int. Conf. Neural Networks,* 1997, vol. 4, pp. 2507–2512.

[34] A. Rosenfeld, R. A. Hummel, and S. W. Zucker, "Scene labeling by relaxation operations," *IEEE Trans. Syst., Man, Cybern.,* vol. 6, pp. 420–433, 1976.

[35] W. Singer and C. M. Gray, "Visual feature integration and the temporal correlation hypothesis," *Annu. Rev. Neurosci.,* vol. 18, pp. 555–586, 1995.

[36] M. Stoecker, H. J. Reitboeck, and R. Eckhorn, "A neural network for scene segmentation by temporal coding," *Neurocomput.,* vol. 11, pp. 123–134, 1996.

[37] D. Terman and D. L. Wang, "Global competition and local cooperation in a network of neural oscillators," *Physica D,* vol. 81, pp. 148–176, 1995.

[38] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems.* Washington, D.C.: Winston, 1977.

[39] B. van der Pol, "On 'relaxation oscillations'," *Philosophical Mag.,* vol. 2, no. 11, pp. 978–992, 1926.

[40] B. C. Vemuri, A. Mitiche, and J. K. Aggarwal, "Curvature-based representation of objects from range data," *Image Vision Comput.,* vol. 4, no. 2, pp. 107–114, 1986.

[41] C. von der Malsburg, "The correlation theory of brain function," Max-Planck-Inst. Biophys. Chemistry, Int. Rep. 81-2, 1981.

[42] D. L. Wang and D. Terman, "Locally excitatory globally inhibitory oscillator networks," *IEEE Trans. Neural Networks,* vol. 6, pp. 283–286, 1995.

[43] D. L. Wang and D. Terman, "Image segmentation based on oscillatory correlation," *Neural Comput.,* vol. 9, pp. 805–836, 1997. For errata see *Neural Comput.,* vol. 9, pp. 1623–1626, 1997.

[44] M. A. Wani and B. G. Batchelor, "Edge-region-based segmentation of range images," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. 16, pp. 314–319, 1994.

[45] R. Yagel, D. Cohen, and A. Kaufman, "Context sensitive normal estimation for volume imaging," in *Scientific Visualization of Physical Phenomena,* N. M. Patrikalakis, Ed. New York: Springer-Verlag, 1991, pp. 211–234.

[46] S. W. Zucker, "Region growing: Childhood and adolescence," *Comput. Graphics Image Processing,* vol. 5, pp. 382–399, 1976.

Xiuwen Liu (S'97) received the B.Eng. degree in computer science in 1989 from Tsinghua University, Beijing, China, and the M.S. degrees in geodetic science and surveying in 1995 and computer and information science in 1996 both from the Ohio State University, Columbus, OH. He is currently a Ph.D. candidate in the Department of Computer and Information Science at the Ohio State University.

From August 1989 to February 1993, he was with Department of Computer Science and Technology at Tsinghua University. From September 1994 to September 1998, he was with Center for Mapping at the Ohio State University working on various satellite image understanding projects. His current research interests include image segmentation, neural networks, computational perception, and pattern recognition with applications in automated feature extraction and map revision.

Mr. Liu is a member of the American Congress on Surveying and Mapping and the American Society for Photogrammetry and Remote Sensing.

DeLiang L. Wang (M'94–A'95) was born in Anhui, China, in 1963. He received the B.Sc. degree in 1983 and the M.Sc. degree in 1986 from Peking (Beijing) University, Beijing, China, and the Ph.D. degree in 1991 from the University of Southern California, Los Angeles, CA, all in computer science.

From July 1986 to December 1987 he was with the Institute of Computing Technology, Academia Sinica, Beijing. Since 1991, he has been with Department of Computer and Information Science and Center for Cognitive Science at the Ohio State University, Columbus, OH, where he is currently an Associate Professor. His present research interests include neural networks for perception, neurodynamics, neuroengineering, and computational neuroscience.

Dr. Wang is a member of AAAS, the IEEE Computer Society, and the International Neural Network Society. He is a recipient of the 1996 U.S. Office of Naval Research Young Investigator Award.