# Motion Segmentation Using Temporal Block Matching and LEGION

Erdogan Çesmeli[1], Deliang L. Wang[2], Delwin T. Lindsey[3], and James T. Todd[3]

[1]Center for Biomedical Engineering
[2]Department of Computer and Information Science and Center for Cognitive Science
[3]Department of Psychology and Center for Cognitive Science
{cesmeli, dwang}@cis.ohio-state.edu
{dlindsey, jtodd}@magnus.acs.ohio-state.edu
The Ohio State University, Columbus, OH 43210, USA

## Abstract

*A motion segmentation method is proposed for an input sequence of random dot and binary images. The method is composed of two main stages, inspired by primate visual system. The first stage determines local velocity information at each location in every image frame using its two neighboring image frames. Measurements of a particular velocity at all locations form the corresponding velocity layer. The second stage performs segmentation based on the motion information in the velocity layers. Each velocity layer provides input to a LEGION (Locally Excitatory Globally Inhibitory Oscillator Networks), which is a 2D array of neural oscillators. When LEGION networks are simulated, the oscillators corresponding to the region of a uniform velocity oscillate in synchrony, whereas the regions with different velocities tend to attain different phases. Final output is displayed in the segmentation network. Results demonstrating the performance of our method on synthetic image sequences are provided and related to psychophysics.*

## 1. Introduction

Motion is a ubiquitous property of real visual scenes. Visual motion can be defined as the changes of luminance over time throughout the visual field. Presence of motion in a scene increases the complexity of visual analysis but it also provides extra information for segmentation. Motion information is a major cue to identify camouflaged targets and sometimes, is the only cue for figure/ground segregation. It is one of the Gestalt grouping principles, which is also known as common fate [6]. Thus, a computational vision model involving dynamic scene analysis has to have motion as one of its functional tokens.

Approaches to motion analysis can be categorized into two groups. The first group includes approaches based on tracking salient features in the frames of an image sequence (see [10] for review). The second group involves motion energy filters and seeks representation of moving patterns in the spatiotemporal frequency domain [1], [10], [16]. Gradient based approaches fall under the second category since the derivative operation can also be viewed as one type of filtering [11]. The Elementary Motion Detector (EMD) developed by Reichardt [3] with appropriate addition of spatial and temporal filters [9] is considered to be equivalent to motion energy models. In its simple form, EMD falls under the first category as well, since it involves correlation of luminance at different locations across time.

Our method consists of two stages. The first stage extracts motion information by performing temporal block matching (TBM), where intensities within displaced blocks are correlated across consecutive frames. By evaluating TBM at all locations a stack of velocity layers is obtained. In the second stage, each velocity layer is associated with a corresponding LEGION network [13]. LEGION has been proposed to deal with static image segmentation [15]. It is based on the idea of oscillatory correlation [7], [13], whereby phases of neural oscillators encode region labeling. Oscillators corresponding to one region have the same phase which is different from that of other regions [15]. In this paper, we describe a method that combines TBM and LEGION for motion segmentation.

## 2. Model Components

### 2.1 Temporal block matching (TBM)

EMD, which is a motion model based on fly's visual system, is composed of two symmetric parts including the rightward and the leftward motion detectors. Rightward motion is detected by correlating the pixel intensity at the location *(x, t)* with that of *(x+*$\Delta x$*, t+*$\Delta t$*)*. Here, the model implicitly assumes a speed of $\Delta x/\Delta t$. Leftward motion detector does the same operation with *(x, t)* and *(x-*$\Delta x$*,*

0-7803-4859-1/98 $10.00©1998 IEEE

$t+\Delta t$). The difference between the two correlations yields the final output. Depending on the sign of the output, motion of the pixel is inferred to be rightward or leftward.

Unlike EMD, TBM tries to match intensities within a local block (window) of size $w \times w$ instead of a single pixel. Selectivity to different velocities is achieved by sliding the correlation windows by different quantities and in different directions on the consecutive frames as shown in Figure 1. For TBM employing three consecutive image frames, correlation corresponding to velocity ***r*** at location $j$ and time $t$ can be expressed:

$$v_{rj} = \sum_{i=1}^{\|W\|} F_{t-1}\left(W_{-d_r}(i)\right) F_t(W_0(i))\, F_{t+1}\left(W_{+d_r}(i)\right) \quad (1)$$

where $F_t$ is the $t^{th}$ image frame, $W_d(i)$ is the $i^{th}$ element of the window whose center location is shifted by a displacement vector $\boldsymbol{d}$ with respect to location $j$. There is a total number of $R$ different velocities in the model.

By evaluating correlations for a particular displacement vector at all locations, a velocity layer is formed. Different displacement vectors lead to different layers of velocity detectors. Velocity detectors corresponding to the same location form a velocity column. For a pixel in a uniformly moving region, one velocity detector is expected to have the strongest output among the others in its velocity column. For a pixel close to a motion boundary or in a region of transparent motion, more than one velocity detector might attain high value in its velocity column. For a TBM employing three consecutive frames, there is a stack of velocity layers for each image frame except for the first and the last frames in the input image sequence (Fig. 2)

## 2.2 LEGION

LEGION is based on the idea of oscillatory correlation, where the phases of the neural oscillators encode the binding of the features [13], [14]. A single oscillator $i$ of LEGION is defined as a feedback loop between an excitatory unit $x_i$ and an inhibitory unit $y_i$:

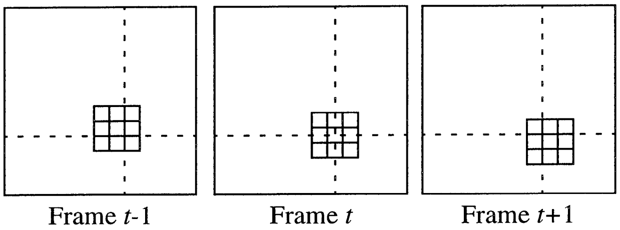

**Figure 1. Correlation calculation for a velocity of 1 pixel/frame to the right and bottom. Corresponding elements of the three consecutive local windows are multiplied and summated to obtain the correlation.**

$$\frac{dx_i}{dt} = 3x_i - x_i^3 + 2 - y_i + S_i + \rho \quad (2a)$$

$$\frac{dy_i}{dt} = \varepsilon\left(\gamma\left(1 + \tanh\left(\frac{x_i}{\beta}\right)\right) - y_i\right) \quad (2b)$$

where $S_i$ is coupling, $\rho$ denotes the variance of a Gaussian noise term, and $\gamma$ and $\beta$ are system parameters. The x-nullcline of (2a) is a cubic curve while the y-nullcline of (2b) is a sigmoid function, as shown in Figure 3. If these curves intersect along the middle branch of the cubic nullcline, then the system is oscillatory. If the nullclines intersect at a point along the left branch of the cubic (a stable fixed point), the system does not oscillate. The parameter $\varepsilon$ is chosen to be small positive number so that (2) defines a relaxation oscillator with two time scales. In the limit cycle, oscillator $i$ travels along the left branch (LB) and jumps to the right branch (RB) and thus, becomes active. Once it reaches the right knee (RK) it jumps back to LB completing the limit cycle.

$S_i$ denotes coupling from other oscillators and global inhibitor (GI), and the external input $I_i$ in the network:

$$S_i = I_i + \sum_{k \in N(i)} W_{ik}\, H(x_k - \theta_x) + W_p\, H(p_i - \theta) - W_z\, H(z - \theta_{xz}) \quad (3)$$

where $W_{ik}$ is the connection weight from oscillator $k$ to

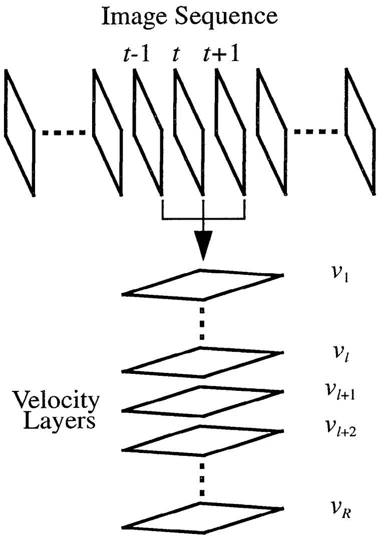

**Figure 2. Corresponding to each image frame $F_t$, there is a stack of velocity layers.**

oscillator $i$, $H$ is the Heaviside step function, $W_p$ is the weight for potential, and $N(i)$ represents the neighborhood topology in the array (Fig. 4). A variable called potential is introduced for each oscillator to make distinction between homogeneously stimulated and noisy regions:

$$p_i' = \lambda(1-p_i)H\left[\sum_{k\in N_p(i)} H(x_k-\theta_x)-\theta_p\right]-\mu p_i \qquad (4)$$

where $\lambda > 0$, $\mu$ is in the order of $\varepsilon$, and $N_p$ is the potential neighborhood, which is larger than $N$. Initially, oscillators have high potentials which continuously decay. When an active oscillator has a number of active neighbors more than $\theta_p$ in its $N_p$, its potential rises to 1, otherwise it is reduced to 0. Oscillators that maintain high potential are called leaders while others are called followers. Large homogeneous regions can produce leaders. Since noisy fragments tend to be small and isolated, they tend not to be able to produce leaders. When a LEGION network runs, groups of oscillators that receive similar stimuli and correspond to topologically connected regions form a segment. Only a leader can start formation of a segment. Segments lacking leaders cannot become active and will stop oscillating as in the case of noisy fragments. $\theta_x$, $\theta$, and $\theta_{xz}$ are thresholds and $W_z$ is the weight of the inhibitory connection from GI, whose activity, $z$, is defined as

$$\frac{dz}{dt} = \varphi(\sigma_\infty - z) \qquad (5)$$

where $\sigma_\infty = 1$ if $x_i \geq \theta_{zx}$ for at least one oscillator and $\sigma_\infty = 0$ otherwise, and $\varphi$ is a constant. The typical neural network structure used in our image analysis is a two dimensional array of oscillators and one GI as shown in Figure 4. LEGION has been applied to static image segmentation [15].

## 3. The Model: TBM and LEGION

There is a strong evidence from both psychology and neurophysiology that there exist at least two separate stages of motion processing in primate visual system (see [10] for review). Similarly, our method is composed of two stages including TBM and LEGION networks.

Even though TBM in our method is similar to the local motion extraction method in [2], we employ two neighboring frames instead of just one (Fig. 5). Because of temporal locality, segmentation result for a given image frame is not affected by temporally distant frames. For a given image sequence, one segmentation result is determined for each image frame except for the first and the last frames. Prior to TBM, input sequence of images is filtered by a Laplacian of Gaussian filter (LoG) of size $m \times m$ [8]. Subsequently, filtered images are thresholded with 0 to make them binary. Following that, a set of correlators is applied to each group of three consecutive image frames in the sequence. Thus, for each group, a stack of velocity layers is obtained. One can view a velocity layer as the spatial distribution of a measure analogous to the likelihood of a particular velocity. In the second stage of our method, there is one LEGION network corresponding to each velocity layer. Correlations forming a velocity layer define the input to the corresponding LEGION network (Fig. 5). Consequently, the coupling weight between two oscillators, $i$ and $k$, on a velocity layer, $r$, is determined based on their correlations, given by:

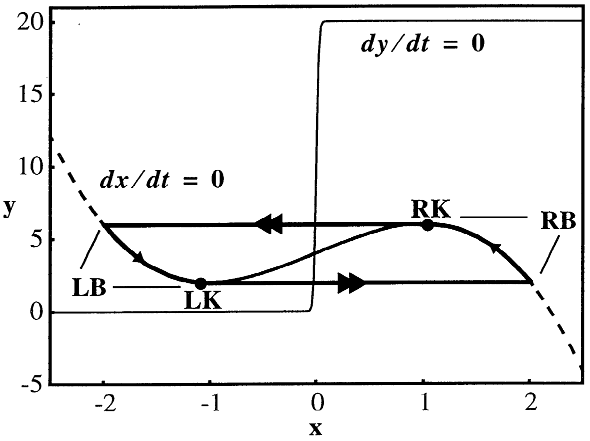

**Figure 3. Phase plane diagram of a single oscillator. The x-nullcline is the dotted line and the y-nullcline is the solid line. Since the nullclines intersect only along the middle branch of the x-nullcline here, the oscillator produces a limit cycle, drawn as thick solid line. The parameters are $\varepsilon = 0.005$, $\gamma = 10.0$, $\beta = 0.02$, $I_i = 2.0$, and $\rho = 0.02$.**

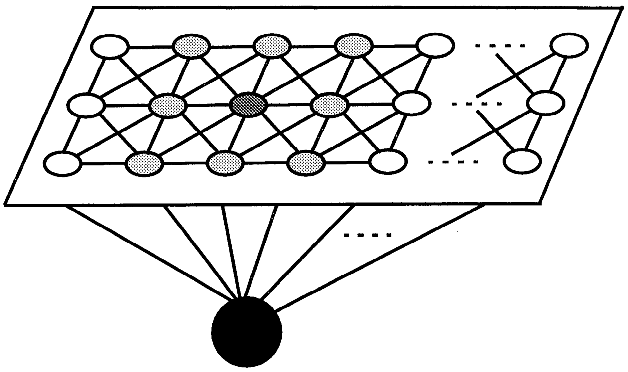

**Figure 4. Architecture of a 2-D LEGION network with 8-Nearest Neighbor coupling. The global inhibitor (GI) is indicated by the black circle.**

$$W_{rik} = \frac{1}{|v_r(i) - v_r(k)|} \tag{6}$$

where the denominator is added a small number to avoid division by zero. Thus, if two neighboring oscillators have similar correlations they will have strong coupling.

In this multilayered architecture of LEGION networks we modified $S_i$ in (3) as $S_{ri} = H(S - \theta_s)$ where $\theta_s$ is a threshold and

$$\begin{aligned} S = {} & \sum_{k \in N(i)} W_{rik}\, H(x_{rk} - \theta_x) - W_z H(z - \theta_{xz}) \\ & + W_p\, H(p_{ri} - \theta)\, H\left[1 + \sum_{c=1}^{R} H(v_{ri} - v_{ci}) - R\right] \\ & + I_i H\left[1 + \sum_{j \in N_p(k)} I_j - \|N_p\|\right] \end{aligned} \tag{7}$$

There is an inhibitory interaction within each velocity column through the potential term. This interaction is crucial when there is more than one leader in a velocity column. As a result of this interaction, the oscillator that has the largest input becomes the winner and thus, can start forming its segment on its velocity layer. Other oscillators on the same velocity layer are recruited through local couplings. Oscillators in a segment become active simultaneously (synchronization). Because of the global inhibition, during the active phase of a segment, no other segment can be formed (desynchronization). If a leader within a velocity column is not a winner, it can only be recruited by some other leader(s) on its velocity layer. Therefore, more than one oscillator in a velocity layer can become active and thus, different velocities can be represented at the same location. This ability has a key role in the representation of transparent motion.

By the addition of the last term in (7), oscillators that correspond to regions with uniform luminances can also become active. However, leaders in these regions cannot become winner. Because of filtering by LoG and thresholding, regions with uniform luminances can only have zero velocity. Because of the inhibition in velocity columns, they cannot be winner but can only be followers.

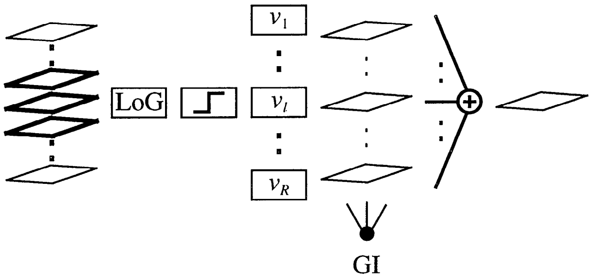

**Figure 5. Flow diagram of the method.**

Finally, a segmentation network is employed where each unit has the summated activity of the corresponding velocity column. Since there is always one active segment at any time, the segmentation network displays the segments in the order they become active.

## 4. Results

To demonstrate the performance of our method, a set of synthetic image sequences is prepared. We employed a set of random dot images because they are widely used in psychophysical experiments and are able to induce several natural visual motion. In these images, segmentation of individual frames is not possible based on a static image analysis. Thus, motion information is the only cue for segmentation. We also included a binary image sequence with a moving region of uniform luminance to test the performance of our method on the challenging blank-wall problem [11].

In all random dot image sequences used in this study, the pattern of dots forming the object(s) does not change throughout the sequence. A new set of random dots corresponding to the region outside the object(s) is generated independently at each time frame. Prior to motion extraction, images are filtered by LoG of size $5 \times 5$ with $\sigma = 0.05$ and thresholded with 0. There are five different velocity layers included in TBM, where a block of size $5 \times 5$ is employed for correlation. These velocities are (0,0), (1,0), (0,1), (-1,0), and (0,-1) pixel/frame, where the values represent horizontal and vertical components of the velocities, respectively. In the simulation of LEGION networks, a functionally equivalent algorithm is employed, where only the following parameters are needed [15]: $N = 3 \times 3$, $N_p = 7 \times 7$, $W_z = 0.74$, $W_z = 1$, $\theta_p = 36.75$, and $I_i = 1$. The only parameter that needs to be adjusted from one image sequence to another is $\theta_s$. All images and LEGION networks are of size $64 \times 64$.

We illustrate the segmentation process using the example shown in Figure 6, where two square regions are moving oppositely with a horizontal speed of 1 pixel/frame. A schematic diagram of the input is depicted in Figure 6A. Since TBM requires only three consecutive frames in the calculations, we demonstrate the method for only one group of three frames (Fig. 6B). Segmentation results corresponding to other groups can be obtained similarly. In our method, first, images are filtered with LoG and thresholded (Fig. 6C). As a result of TBM, five velocity layers are obtained as depicted in Figure 6D. Except for spurious responses in other velocity layers, each square is detected only by the velocity layer that correctly corresponds to its motion. Subsequently, five LEGION networks are simulated with $\theta_s = 0.4$ and random initial conditions. Consequently, the final result is read out from the segmentation network as shown in Figure 6E. The first frame shows the initial random activities of the segmentation network. After a few cycles, a segment is formed on the rightward velocity layer. Activities of this segment are reflected onto the segmentation network as depicted in the second frame. Subsequently, oscillators on the leftward velocity layer become active. Corresponding activity of the segmentation network is shown in the third frame. These two regions become active repeatedly during the entire simulation. Because the dots forming the squares are sparse so are the synchronized oscillators. Since oscillators outside moving regions have neither uniform luminance input nor uniform input from velocity layers, they always stay silent and thus, form the background.

Four additional segmentation examples are provided to show the performance of our method on a widely used set of stimulus in psychophysical experiments (Figs. 7-10). For illustration purposes, only schematic diagrams of the inputs are shown when the actual input is a random dot image. In the first example, the input image is divided into two regions in terms of motion. The upper half is moving to the left while the lower half is moving to the right with the same speed of 1 pixel/frame (Fig. 7A). After running LEGION networks with $\theta_s = 0.4$, the segmentation network reaches the periodic activity where oscillators corresponding to each half synchronized (Fig. 7B and C). The second and the third examples illustrate the performance of our method in the presence of transparent motion. In the second example, there are two oppositely moving interleaved regions that are covering the entire image (Figure 8A). They are moving with a vertical speed of 1 pixel/frame. When $\theta_s = 0.3$, the segmentation network shows that there are two moving surfaces covering the entire image in accordance with perception (Figs. 8B and C). The third example is a combination of the examples in Figures 6 and 8. Two rectangular regions, which partially and transparently overlap, are moving oppositely with a horizontal speed of 1 pixel/frame (Fig. 9A). When $\theta_s = 0.4$, the segmentation network shows that there are two rectangular regions. Furthermore, the common region becomes active with both regions as shown in Figures 9B and C. In the final sequence, a uniformly white rectangular region is moving to the right with a speed of 1 pixel/frame (Fig. 10 A). This image has the blank-wall problem, where the true motion of the rectangle's inner region is difficult to recover [11]. In spite of this, by the help of the last term in (7), our method is able to segment the inner region together with the edges whose motion is detected correctly (Fig. 10 B). Segmentation result is obtained with $\theta_s = 0.4$.

## 5. Conclusion

Inspired by primate visual system, our method is composed of two stages, which are TBM and LEGION. Both stages are based on the notion of local computations, which is important for parallel and distributed processing.

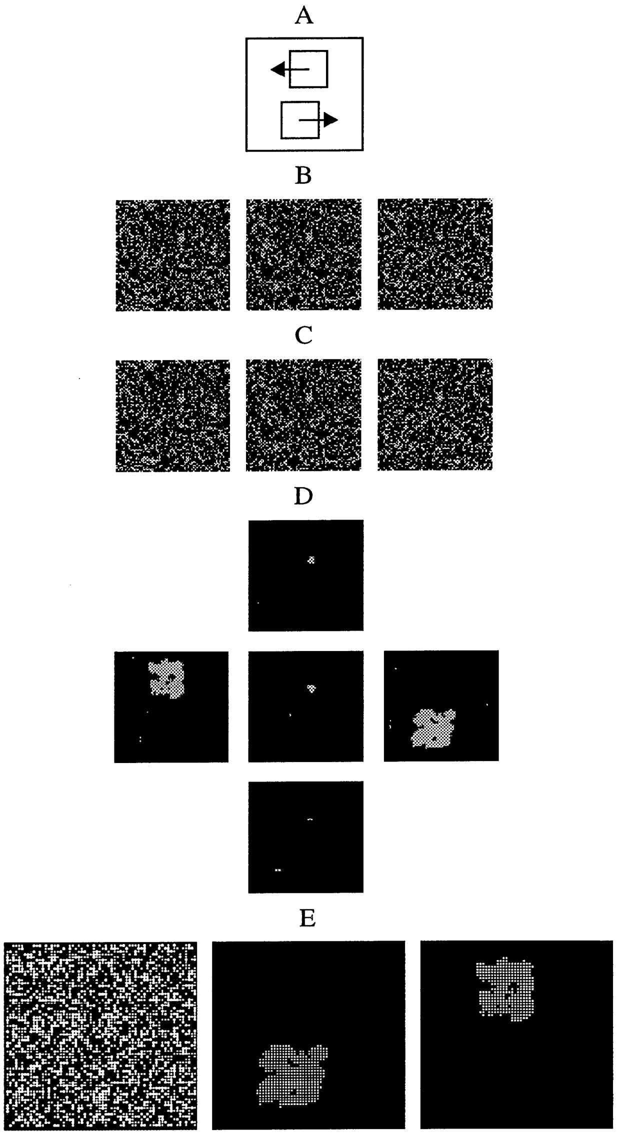

**Figure 6. A) A schematic diagram of the input. B) The input image sequence. C) After LoG and thresholding. D) Velocity layers, stationary, rightward, leftward, upward, and downward, organized in consistent with their directions. E) Initial and repeated activities in the segmentation network.**

Even though our examples are composed of random dot images, which are relatively simpler than real images, they capture many properties of visual motion. Current implementation of our method has only a limited number of velocity layers, which can be increased easily without

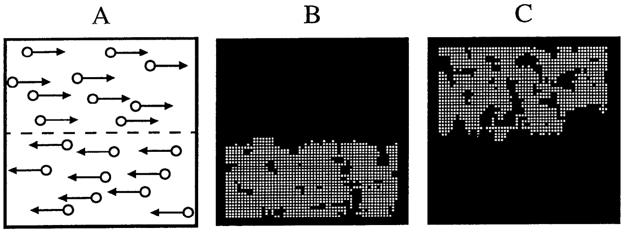

Figure 7. A) A schematic diagram of the input. B-C) Periodic activities in the segmentation network.

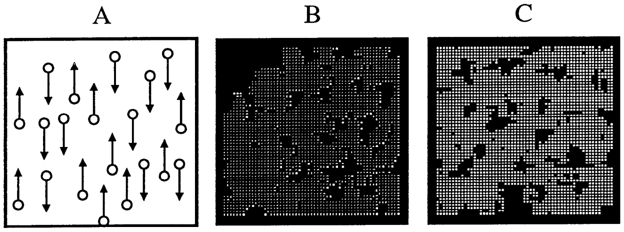

Figure 8. A) A schematic diagram of the input, where two interleaved surfaces are oppositely moving. B-C) Periodic activities in the segmentation network.

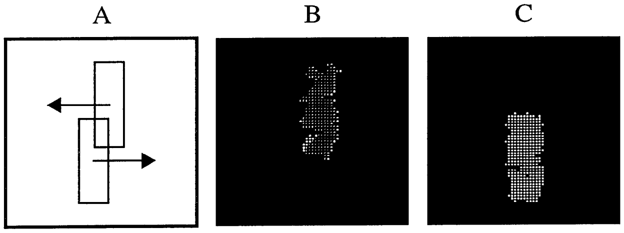

Figure 9. A) A schematic diagram of the input, where two oppositely moving rectangles partially and transparently overlap. B-C) Periodic activities in the segmentation network. The overlapping region is grouped with both rectangles.

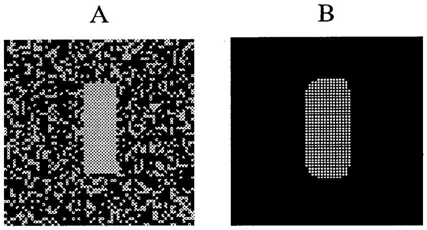

Figure 10. A) The input image frame. B) Periodic activity in the segmentation network.

altering the architecture. Our method has the potential of multiscale analysis by employing a set of LoG filters with different scales. As compared to other neural networks for motion analysis, our model is able to handle more units efficiently [4], [5], [17]. Future goals include testing our method on real images as well as on other psychophysical characteristics of motion perception such as direction repulsion [4].

## Acknowledgments

This work was supported in part by an ONR grant (N00014-93-10335), an NSF grant (IRI-9423312), an ONR Young Investigator Award (N00014-96-1-0676) to DLW, and an OSU Interdisciplinary Seed Grant.

## References

[1] E. H. Adelson and J. Bergen, "Spatiotemporal energy models for the perception of motion," *J. Opt. Soc. Amer. A*, 2:284-299, 1985.

[2] H. Bülthof, J. Little, and T. Poggio, "A parallel algorithm for real-time computation of optical flow," *Nature*, 337:549-553, 1989.

[3] B. Hassenstein and W. E. Reichardt, "Functional structure of a mechanism of perception of optical movement," *Proc. First International Congress of Cybernetics in Namar*, 797-801, 1956.

[4] E. Hirisi and R. Blake, "Direction repulsion in motion transparency," *Visual Neuroscience*, 13:187-197, 1996.

[5] M. Kikuchi and K. Fukushima, "Neural network model of visual system: binding form and motion," *Neural Networks*, 9:1417-1427, 1996.

[6] K. Koffka, *Principles of Gestalt Psychology.* New York: Harcourt, 1935.

[7] C. von der Malsburg and W. Schneider, "A neural cocktail-party processor," *Biol. Cybern.*, 54:29-40, 1986.

[8] D. Marr and E. C. Hildreth, "Theory of edge detection," *Proc. Royal Society of London, B*, 207:187-217, 1980.

[9] J. P. H. van Santen and G. Sperling, "Elaborated Reichardt Detectors," *J. Opt. Soc. Amer. A*, 2:300-321, 1985.

[10] M. E. Sereno, *Neural Computation of pattern motion: modeling stages of motion analysis in the primate visual cortex.* Cambridge MA: MIT press, 1993.

[11] E. P. Simoncelli, Distributed Analysis and Representation of Visual Motion, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge MA, 1993.

[12] O. Sporns, G. Tononi, and G. M. Edelman, "Modeling perceptual grouping and figure-ground segregation by means of active reentrant connections," *Proc. Natl. Acad. Sci. USA*, 88:129-133, 1991.

[13] D. Terman and D. L. Wang, "Global competition and local cooperation in a network of neural oscillators," *Physica D*, 81:148-176, 1995.

[14] D. L. Wang and D. Terman, "Locally excitatory globally inhibitory oscillator networks," *IEEE Trans. Neural Networks*, 6:283-286, 1995.

[15] D. L. Wang and D. Terman, "Image segmentation based on oscillatory correlation," *Neural Comp.*, 9:805-836, 1997.

[16] A. B. Watson and A. J. Ahumada, "Model of visual-motion sensing," *J. Opt. Soc. Amer. A*, 2:322-341, 1985.

[17] H. R. Wilson and J. Kim, "A model for motion coherence and transparency," *Visual Neuroscience*, 11:1205-1220, 1994.