Object Pattern Recognition Below Clutter in Images

Robert Linnehan, John Schindler, Anteon Corporation, Hanscom AFB, MA
E-mail: robert.linnehan2@hanscom.af.mil, Tel: (781)377-1732, Fax: (781)377-8984
Leonid Perlovsky, 2LT Chris Mutz, Air Force Research Laboratory,
Sensors Directorate, Hanscom AFB, MA
Roger Brockett, Harvard University, Cambridge, MA

Abstract—*We are developing a technique for recognizing patterns below clutter based on modeling field theory. The presentation briefly summarizes the difficulties related to the combinatorial complexity of computations, and analyzes the fundamental limitations of existing algorithms such as Multiple Hypothesis Testing. A new concept, dynamic logic, is introduced along with an algorithm suitable for pattern recognition in images with intense clutter data. This new mathematical technique is inspired by the analysis of biological systems, like the human brain, which combines conceptual understanding with emotional evaluation and overcomes the combinatorial complexity of model-based techniques. The presentation provides examples of object pattern recognition below clutter.*

1. Introduction

The algorithm proposed in this paper is developed for a specific solution but has broad applications. Relating data to conceptual content is the challenge of many current methodologies in data mining, sensor\data fusion, signal processing and exploitation. We use modeling field theory (MFT) [1] to relate data and object-concepts. Just as the human mind relates internal visual representations of objects with retinal signals, MFT associates a moving object-concept model with the input image. Well-known model-based techniques such as multiple hypothesis testing (MHT) rely on trying many combinations of models and data to find the object model that best fits the data. MHT is governed by formal logic, which starts with exact statements and derives exact conclusions. But the human mind relates models to data at a pre-conscious, fuzzy level. The mind does not sift through every possible internal visual object-model, in every possible orientation, and in combination with other objects to recognize an object, such as a car on the road. Rather visual models are associated pre-consciously with retinal signals to quickly accomplish object recognition. In a similar manner to visual object recognition, MFT associates models and data "pre-consciously". The algorithm that accomplishes MFT is referred to as dynamic logic [1]. Whereas MHT relies on formal logic, MFT relies on dynamic logic, which starts with fuzzy statements and derives exact conclusions.

Dynamic logic mitigates combinatorial complexity by comparing all models and data simultaneously. Fuzzy models dynamically converge onto corresponding data using feedback from a similarity measure. Each model can then be understood as behaving like an intelligent agent by refining its parameters to extract maximum object information from the data. Just like internal visual models "grab" onto the corresponding internal retinal signals, internal models of stationary and moving objects converge onto patterns in images that best fit the models. The computational cost of the dynamic logic algorithm (DLA) increases linearly with the number of object models, while computations of the classical MHT algorithm increase exponentially with additional models.

Problem Statement

A common problem with object recognition is the existence of overwhelming clutter data in the image. Figure 1 displays the image of an object moving in a straight line without the addition of clutter. When clutter is added to the image in Figure 2, the object becomes much more difficult to observe.

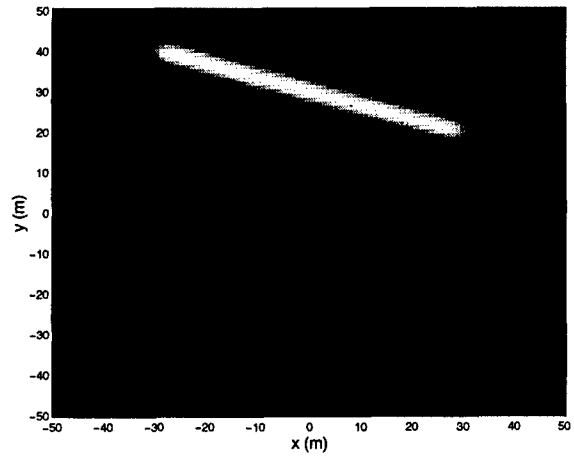

Fig. 1. Object signal

The signal to clutter ratio (SCR) is an indication of the difficulty to recognize an object pattern in an image. The SCR is defined as the ratio of the maximum absolute

KIMAS 2003, October 1-3, 2003, Boston, MA, USA.

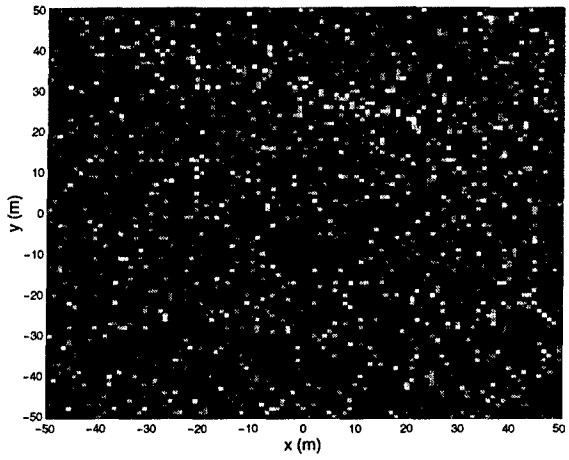

Fig. 2. Object signal + clutter

value of the object signal over the average clutter power. The iterative recognition algorithm we present is able to positively determine if the object is present in an image, even for very low SCR (such as Figure 2 with SCR = 2.4 dB).

Models

We utilize modeling field theory (MFT) to develop an iterative DLA that is able to determine the presence of an object signal in an image buried in clutter. A number of models are introduced that account for the presence of every pixel in the image. The models are probability distribution functions: $\mathrm{pdf}(\mathbf{n}|\mathbf{m})$, for the signal in pixel n given it came from model m. Clutter is one model that we index with m = 1. The clutter model (pdf) is uniformly distributed and expressed as:

$$\mathrm{pdf}(\mathbf{n}|1) = \hat{\mathrm{r}}_1. \tag{1}$$

The parameter $\hat{\mathrm{r}}_1$ is a single number that represents the estimated proportion of signal power in the entire image coming from clutter.

The moving object model is indexed by m = 2. The model (pdf) for all pixels n given the signal of a particular pixel came from the object is expressed as:

$$\mathrm{pdf}(\mathbf{n}|2) = \frac{1}{\mathrm{T}} \sum_{\mathbf{t}} \hat{\mathrm{r}}_2 \cdot \mathrm{G}(\mathbf{n}|\hat{\mathrm{x}}_\mathrm{t}, \sigma_\mathrm{x}), \tag{2}$$

where the parameter $\hat{\mathrm{r}}_2$ represents the proportion of signal power coming from the object. The pdf for the object model is computed using a sum of Gaussian distributions G over time t = 1, 2,..., T. The parameter $\hat{\mathrm{x}}_\mathrm{t}$ represents the estimated position of the object over t in both the x and y dimensions. For instance, if the moving object model is a straight line, than the linear equations,

$$\begin{aligned} \mathbf{x} &= \hat{\mathrm{x}}_0 + \hat{\mathrm{v}}_\mathrm{x}\mathbf{t}, \\ \mathbf{y} &= \hat{\mathrm{y}}_0 + \hat{\mathrm{v}}_\mathrm{y}\mathbf{t}, \end{aligned} \tag{3}$$

describe the object's position, where $(\hat{\mathrm{x}}_0, \hat{\mathrm{y}}_0)$ are the object's estimated initial component positions and $(\hat{\mathrm{v}}_\mathrm{x}, \hat{\mathrm{v}}_\mathrm{y})$ are its estimated component velocities. The parameter σ_x represents the standard deviation of the Gaussian distributions, also in both the x and y dimensions. The number of terms in the summation (eq. 2) should be sufficient to accurately model the entire moving object pattern.
The total pdf for all pixels is the summations of the conditional pdfs:

$$\mathrm{pdf}(\mathbf{n}) = \mathrm{pdf}(\mathbf{n}|1) + \mathrm{pdf}(\mathbf{n}|2). \tag{4}$$

It is not required that the object pattern be a straight line as in the example above. The pattern can exhibit any curvature or shape and be modeled by a non-linear equation.

2. The Dynamic Logic Algorithm

The procedure of the DLA is summarized as follows [1], [2]:

1) Set initial value estimates for the unknown object parameters $\hat{\mathrm{x}}_0, \hat{\mathrm{y}}_0, \hat{\mathrm{v}}_\mathrm{x}, \hat{\mathrm{v}}_\mathrm{y}, \hat{\mathrm{r}}_2$. Also set initial values for the standard deviations σ_x and σ_y.

In absence of any information about the object, we set the model initial position in the center of the image with zero velocity. Referring to Figure 2, $\hat{\mathrm{x}}_0 = \hat{\mathrm{y}}_0 = 0\mathrm{m}, \hat{\mathrm{v}}_\mathrm{x} = \hat{\mathrm{v}}_\mathrm{y} = 0\mathrm{m/s}$.

The standard deviations are initialized to a large value, including the entire image, corresponding to the uncertainty of knowledge:

$$\sigma_\mathrm{x} = \sigma_\mathrm{y} = \sqrt{\mathrm{N}}\Big/2, \tag{5}$$

where N is the total number of image pixels.
The proportion of clutter power and the proportion of object power are initialized to be:

$$\begin{aligned} \hat{\mathrm{r}}_1 &= \Big[\sqrt{\mathrm{N}} - 1\Big]\Big/\sqrt{\mathrm{N}}, \\ \hat{\mathrm{r}}_2 &= 1 - \hat{\mathrm{r}}_1. \end{aligned} \tag{6}$$

2) With values assigned to the estimated parameters, the probability distribution functions, eq. (1), (2), can be computed, as well as the total $\mathrm{pdf}(\mathbf{n})$.

3) The probabilities for each pixel are then computed as follows:

$$\begin{aligned} \mathrm{P}(\mathbf{n}|1) &= \tfrac{\mathrm{pdf}(\mathbf{n}|1)}{\mathrm{pdf}(\mathbf{n})}, \\ \mathrm{P}(\mathbf{n}|2) &= \tfrac{\mathrm{pdf}(\mathbf{n}|2)}{\mathrm{pdf}(\mathbf{n})}. \end{aligned} \tag{7}$$

This can be interpreted as follows: $\mathrm{P}(\mathbf{n}|1)$ is the probability that pixel n "belongs to the clutter model" or that the signal in a particular pixel originates from clutter; $\mathrm{P}(\mathbf{n}|2)$ is the probability pixel n "belongs to the object model" or that the signal in a particular pixel originates from the object,

for $n = 1, 2, ...N$. $P(n|1) + P(n|2) = 1$ for $n = 1, 2, ...N$, in correspondence with the total probability (for each pixel receiving any signal) being 1.

Note that these probabilities are estimates based on the current object parameter estimates for a particular iteration. Therefore, these quantities are not "real" probabilities, they can be considered as subjective, estimated probabilities.

4) A similarity measure is computed that evaluates the correspondence of the estimated model to the image data $S(n)$. This can be interpreted as mutual information in models about the data [1]:

$$L = \frac{\sum_{n=1}^{N} S(n) \ln [pdf(n)]}{\sum_{n=1}^{N} S(n)}, \tag{8}$$

where $S(n)$ is the absolute value of the complex signal data in pixel n.
Dynamic logic maximizes this function over the five unknown parameters:

$$\max_{\hat{x}_0, \hat{y}_0, \hat{v}_x, \hat{v}_y, \hat{r}_2} (L) \quad . \tag{9}$$

For the linear pattern equations (3) above, the solution to the unknown parameters that incrementally maximize L (for each iteration) is found analytically by taking the partial derivative with respect to each parameter. The following equations are the result:

$$\begin{aligned} &< x_n > -\hat{x}_0 < 1 > + \hat{v}_x < t > = 0, \\ &< x_n t > -\hat{x}_0 < t > + \hat{v}_x < t^2 > = 0, \\ &< y_n > -\hat{y}_0 < 1 > + \hat{v}_y < t > = 0, \\ &< y_n t > -\hat{y}_0 < t > + \hat{v}_y < t^2 > = 0. \end{aligned} \tag{10}$$

Here y_n and x_n are the known position coordinates of each pixel n. The parameter t is the same time vector used in equations (3).

The brackets <> represent the following function on the known parameters:

$$< u > = \frac{\sum_{n=1}^{N} \sum_{t=0}^{N_p - 1} u(n,t) \, |S(n)| \, P(n|2)}{\sum_{n=1}^{N} |S(n)| \, P(n|2)}, \tag{11}$$

for $u = t, t^2, x_n, x_n t, y_n, y_n t$.

5) Now the proportional power parameter estimates are computed using the updated pdfs and the image data:

$$R(m) = \sum_{n=1}^{N} S(n) \frac{pdf(n|m)}{pdf(n)} \tag{12}$$

$$\hat{r}_m = \frac{R(m)}{\sum_{m=1}^{M} R(m)}, \tag{13}$$

for m = 1, 2.

The parameters, $\hat{x}_0, \hat{y}_0, \hat{v}_x, \hat{v}_y, \hat{r}_1, \hat{r}_2$ are improved estimates for the current iteration.

6) The standard deviations σ_x and σ_y are reduced by a factor between 0.93 and 0.99. (For the examples that are to follow we used a factor of 0.96. We did not yet optimize the convergence process. Convergence of this algorithm was proved in [1])

7) The algorithm iterates to step 2. The process is continued until a predefined value (possibly 2 or 3 pixel resolutions) of σ_x and σ_y is realized. (For our examples it took 80 iterations to reach a minimum required standard deviation of 2 pixel resolutions).

Example 1: A Linear Object Pattern

The DLA is applied to a 101 x 101 pixel image that contains an object signal with the parameters: $x_0 = -30\text{m}, y_0 = 42\text{m}, v_x = 1\text{m/s}, v_y = -0.4\text{m/s}$. The object signal alone is seen in Figure 1. The image data that the DLA operates on has clutter at an SCR of 2.4 dB (see Figure 2). There are 21 Gaussian distributions used to model the object pattern.

Figures 3 through 7 illustrate for several iterations the sum of Gaussian distributions over t in both dimensions x and y.

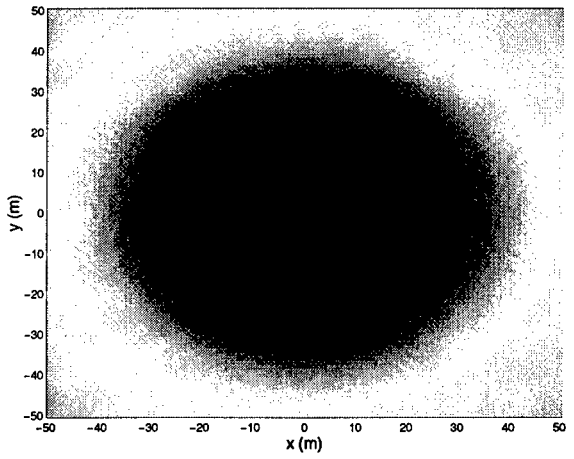

Fig. 3. Iteration 1

Figure 3 shows the first iteration where the initial position estimate is at the center of the image with $\hat{v}_x = \hat{v}_y = 0$, and with standard deviations large enough to include most of the pixels with non-zero probability. Here the Gaussian distributions are located at the same point since the velocity estimates are zero. In Figure 4 we see σ_x and σ_y are reduced

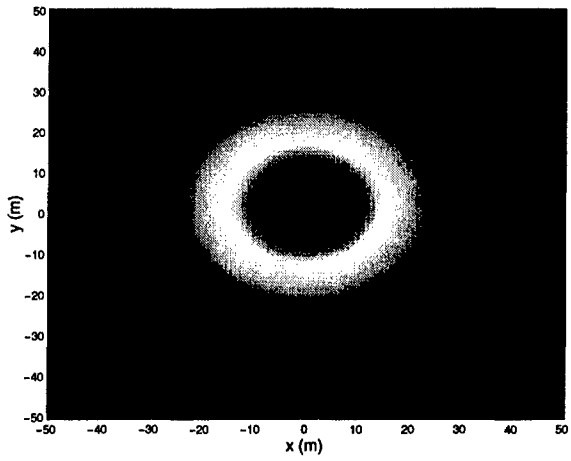

Fig. 4. Iteration 30

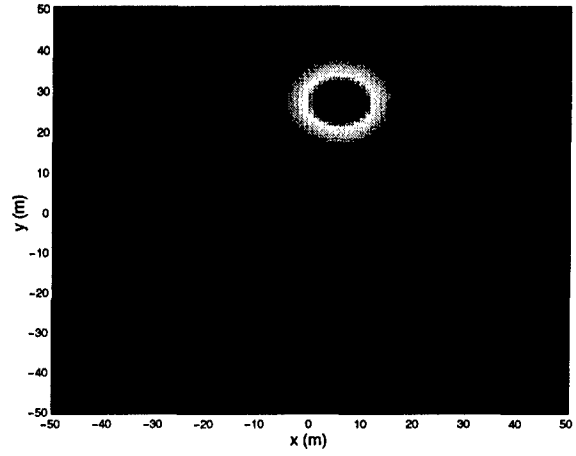

Fig. 5. Iteration 50

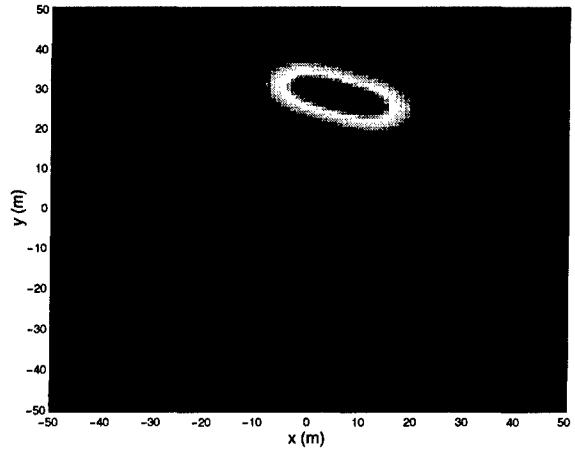

Fig. 6. Iteration 60

but the parameters estimates are still at the image center with no velocity. As σ_x and σ_y are further reduced in Figures 5 and 6, the initial position estimates move near the target and the velocity estimates take on some value. Note the similarity

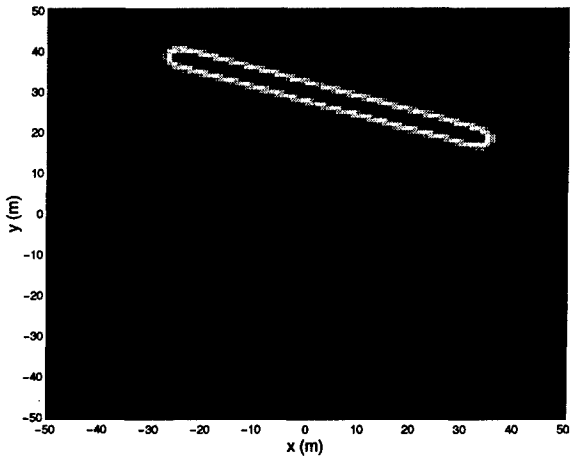

Fig. 7. Iteration 80

in the location and shape of the estimated distributions, Figure 7, to the true object path seen in Figure 1, even though the algorithm was only provided the clutter image in Figure 2.

Multiple Object Recognition

Determining the presence of more than one object requires adding more models to the algorithm. Extending the single object case, m = 1 represents clutter, m = 2 represents one object, m = 3 represents a 2nd object, ... , m = M represents object M-1. Each object has its own parameters to be estimated: $\hat{x}_{0m}, \hat{y}_{0m}, \hat{v}_{xm}, \hat{v}_{ym}, \hat{r}_m$. The conditional pdf for clutter is given above, equation 1, and the pdfs for the remaining object models are computed as:

$$\mathrm{pdf}(\mathbf{n}|\mathbf{m}) = \frac{1}{\mathrm{T}} \sum_{\mathbf{t}} \hat{\mathrm{r}}_{\mathrm{m}} \cdot \mathrm{G}(\mathrm{n}|\hat{\mathrm{x}}_{\mathrm{tm}}, \sigma_{\mathrm{x}}), \qquad (14)$$

for $\mathrm{m} = 2, 3..., \mathrm{M}$. The parameter $\hat{\mathrm{x}}_{\mathrm{tm}}$ is the estimated position of object $\mathbf{m}$ at time t, and $\hat{\mathrm{r}}_{\mathrm{m}}$ is the estimated proportion of signal power coming from object $\mathbf{m}$. The total $\mathrm{pdf}(\mathbf{n})$ is computed as $\sum_{\mathrm{m}} \mathrm{pdf}(\mathrm{n}|\mathrm{m})$ and the similarity measure L is maximized with respect to the unknown parameters for all object models.

Example 2: Three Non-Linear Object Patterns

The following is a case with three objects in the image, and their motion is produces a contoured pattern. Figure 8 displays the input image to the DLA, and Figure 9 displays the object signals without clutter. The signal to clutter ratios for the individual objects are, from the top-right moving clockwise, -0.70 dB, -0.73 dB, and -1.98 dB.

The distributions operating on the clutter image (Figure 8) for the three object models are seen in Figures 10 through 14.

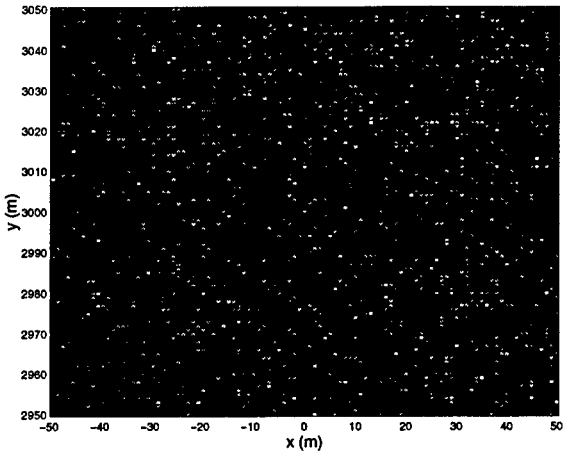

Fig. 8. Object + clutter signals

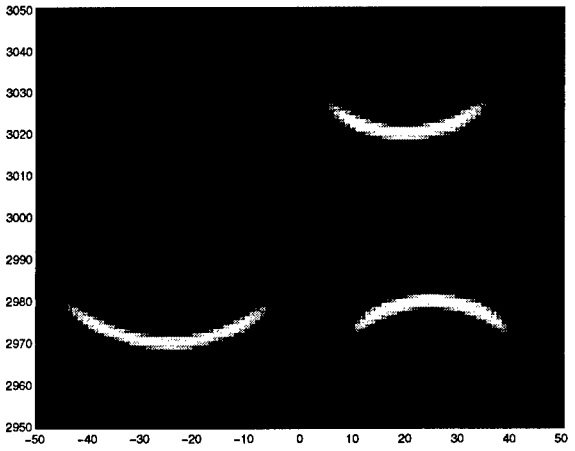

Fig. 9. Object signals

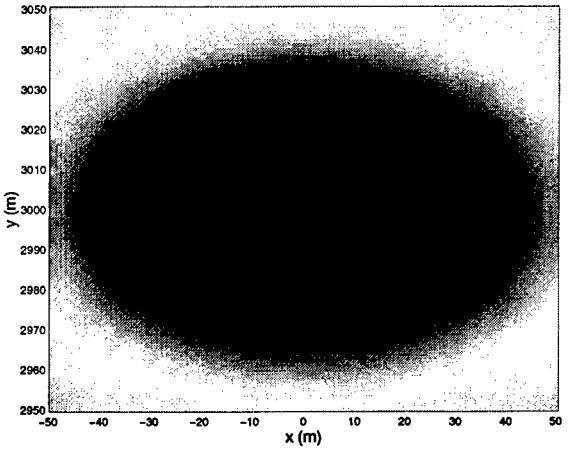

Fig. 10. Iteration 1

Again note the similarity between the final result of the dynamic logic algorithm estimates (Figure 14) and the object signals (Figure 9). The object models were accurately estimated at an SCR where recognition was previously considered unattainable.

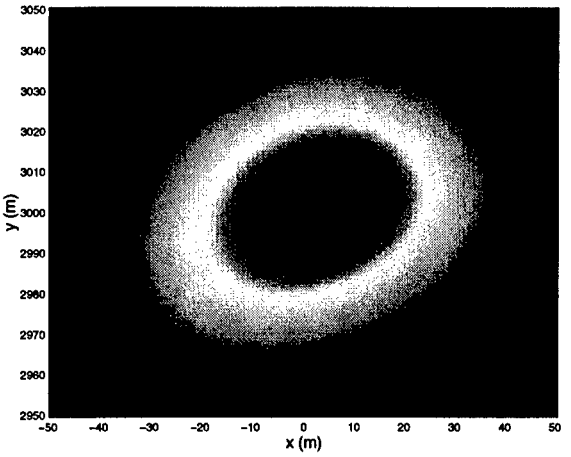

Fig. 11. Iteration 30

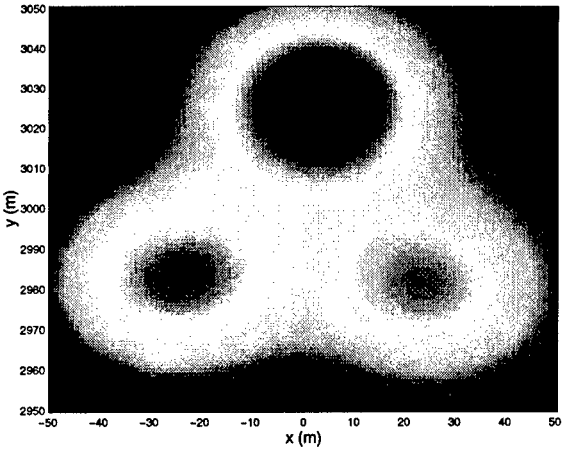

Fig. 12. Iteration 40

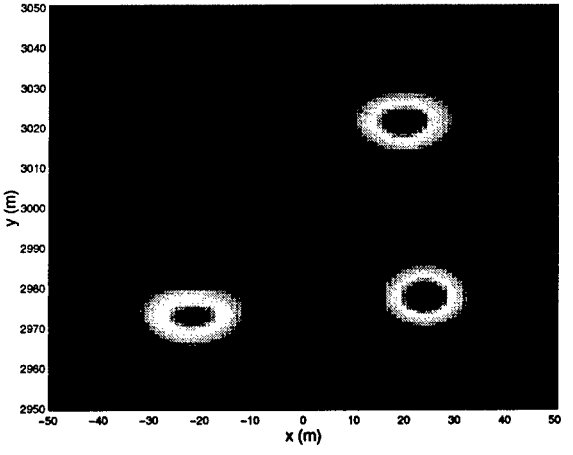

Fig. 13. Iteration 70

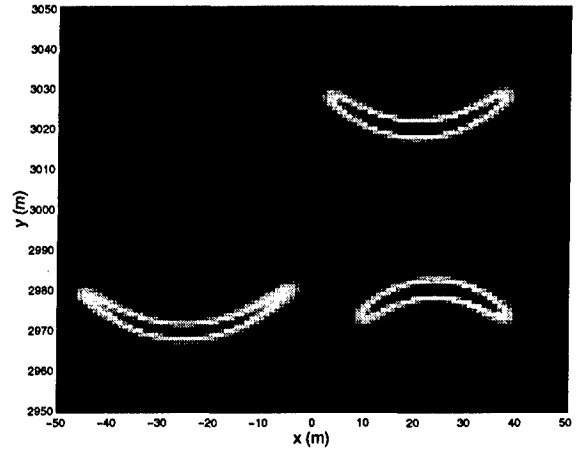

Fig. 14. Iteration 80

3. CONCLUSION

The dynamic logic algorithm (DLA) is a novel technique for model-based estimation and pattern recognition that has proven to overcome previous limitations of model-based techniques. For example, multiple hypothesis testing (MHT) is the general state-of-the-art technique for model-based estimation and recognition restricted by a combinatorial explosion of computations. The addition of object models increases the number of computations exponentially with MHT, whereas a linear increase is expected using the DLA.

Using the DLA on the object patterns illustrated in this paper has yielded an effective performance. We were able to successfully demonstrate an application of the DLA to patterns below clutter, with only a linear increase of complexity for multiple objects.

REFERENCES

[1] Perlovsky, L.I. (2001). Neural Networks and Intellect: using model-based concepts. Oxford University Press, New York, NY.

[2] Perlovsky, L.I., Plum, C.P., Franchi, P.R., Tichovolsky, E.J., Choi, D.S., & Weijers, B. (1997). Einsteinian Neural Network for Spectrum Estimation. Neural Networks, 10(9), pp.1541-46