FINAL REPORT SUBMITTED TO:

National Aeronautics and Space Administration
Lewis Research Center
21000 Brookpark Road
Cleveland, Ohio 44135

IN· 61¬CR

076547

Cooperative Agreement Title:
"Development of Collaborative Research Initiatives to Advance the Aerospace
Sciences - via the Communications, Electronics, Information Systems Focus Group"

Cooperative Agreement Number:
NCC3-320

Submitted by:
Ohio Aerospace Institute
22800 Cedar Point Road
Cleveland, OH 44142
216/962-3000

Principal Investigator:
T. Michael Knasel
Ohio Aerospace Institute
216/962-3040

February 14, 1996

EXECUTIVE SUMMARY

The primary goal of the Adaptive Vision Laboratory Research project was to develop advanced computer vision systems for automatic target recognition. This was collaborative effort between Dr. Louis Tamburino in the Avionics Laboratory at Wright-Patterson Air Force Base and Dr. Mateen Rizki in the Adaptive Vision Laboratory at Wright State University. The approach used in this effort combined several machine learning paradigms including evolutionary learning algorithms, neural networks, and adaptive clustering techniques to develop the E-MORPH system. This system is capable of generating pattern recognition systems to solve a wide variety of complex recognition tasks. A series of simulation experiments were conducted using E-MORPH to solve problems in OCR, military target recognition, industrial inspection, and medical image analysis. The bulk of the funds provided through this grant were used to purchase computer hardware and software to support these computationally intensive simulations. A list of the equipment purchased is shown on page 3. The payoff from this effort is the reduced need for human involvement in the design and implementation of recognition systems. We have shown that the techniques used in E-MORPH are generic and readily transition to other problem domains.

Specifically, E-MORPH is multi-phase evolutionary learning system that evolves cooperative sets of features detectors and combines their response using an adaptive classifier to form a complete pattern recognition system. The system can operate on binary or grayscale images. In our most recent experiments, we used multi-resolution images that are formed by applying a Gabor wavelet transform to a set of grayscale input images. To begin the learning process, candidate chips are extracted from the multi-resolution images to form a training set and a test set. A population of detector sets is randomly initialized to start the evolutionary process. Using a combination of evolutionary programming and genetic algorithms, the feature detectors are enhanced to solve a recognition problem. The design of E-MORPH and recognition results for a complex problem in medical image analysis are described at the end of this report. The specific task involves the identification of vertebrae in x-ray images of human spinal columns. This problem is extremely challenging because the individual vertebra exhibit variation in shape, scale, orientation, and contrast. E-MORPH generated several accurate recognition systems to solve this task. This dual use of this ATR technology clearly demonstrates the flexibility and power of our approach.

Equipment Purchased Under NASA Grant NCC3- 182

Item*	Description	Units Purchased	Cost
DEC Alpha 3000-800	includes expansion chassis, disk drives, monitor, and software	1	$43,074
Gateway 2000 Pentium P60	includes disk drive, monitor, and software	3	$21,668
Zeos Pentium P120	includes disk drive, monitor, and software	1	$5,948
Expansion Disk Drives	includes external and internal units both fixed and removable cartridge systems	8	$5,385
HP LaserJet 4m	Postscript printers	3	$5,196
		TOTAL	**$81,271**

*The first three items include software purchased from the hardware vendor.

E-MORPH: A SYSTEM FOR EVOLVING STRUCTURAL FEATURE DETECTORS FOR AUTOMATIC TARGET RECOGNITION

Mateen M. Rizki
Associate Professor
Department of Computer Science and Engineering
Wright State University

INTRODUCTION

The foundation of a robust pattern recognition system is the set of features used to distinguish among the given patterns. In many problems, the features are predetermined and the task is to build a system to extract the selected features and then classify the resultant measurements. In automatic target recognition problems, the identification of a set of robust, invariant features is complicated because the shape and orientation of the objects of interest are often not known *a priori*. As a result, a human expert is responsible of examining each problem to formulate an effective set of features and then build a system to perform the recognition task. An alternative to this labor intensive approach of building recognition systems has emerged in the past ten years that uses learning algorithms such as neural networks and genetic algorithms to automate the process of feature extraction. There are many advantages to the automated construction of recognition systems over techniques that rely solely on human expertise. Automated approaches are not problem specific. Consequently, once an automated system is developed it can be readily applied to similar problems greatly reducing the time needed to solve new recognition problems. Automated systems are capable of producing solutions that are comparable to the customized solutions created by human experts, but the solutions formed by these systems are often non-intuitive and quite different from the solutions formed by human experts. In many applications, this is a drawback because it is not possible to describe how the solution is obtained. This is also a strength of the automated approach. Automated techniques are unbiased. The features selected to solve problems represent alternative designs based on the structural and statistical attributes of the data. The fact that different features are selected suggests that automated systems are capable of exploring different regions of the space of potential solutions.

E-MORPH [Rizki et. al. 1993, 1994] is an evolutionary learning system that generates pattern recognition systems using several different learning paradigms to automatically perform feature extraction and classification. A robust set of features is identified using a population of pattern recognition systems. Each system is composed of a collection of cooperative feature detectors and a classifier that evolves under the control of a user provided performance measure. The performance measure is typically tied to recognition accuracy, but additional constraints may be included such as complexity measures to sculpt specific types of solutions. The recognition systems compete for survival based on their performance. Successful systems have a higher probability of survival and contribute more information to future generations. The structural and statistical information gathered by each recognition system during the evolutionary

process is passed to the next generation through a process of reproduction with variation. The most successful recognition systems are combined to form new recognition systems that are often superior to either parental unit. Two opposing forces operate in the evolutionary process: exploration and exploitation. By recombining successful solutions during reproduction, each generation contains recognition systems that are more capable of exploiting the performance measure and solving the recognition task. The reproductive process is imperfect, variations in the new recognition systems are created by mutating the structure of the feature detectors. Each new recognition system contains pieces of past successful designs with variations that explore alternative designs. The process of reproduction with variation and selection continues until the best recognition system in the population achieves a satisfactory level of performance.

E-MORPH LEARNING SYSTEM

The overall design of our pattern recognition system is shown in Figure 1. Grayscale images pass through a Gabor wavelet [Gabor, 1946] transformation module that explodes the image into parallel streams of registered images. The Gabor wavelets are oriented bandpass filters that represent an optimal compromise between positional and spatial frequency localization. The Gabor module effectively organizes the spatial frequency and positional information present in the raw grayscale imagery into a registered stack of Gabored images. The target detection module then extracts regions from the full Gabor stack to form registered stacks of small chip-images. The pixel values of each chip are then scaled to fall in the range of -128 to +128. Finally, each stack of chips is processed by the target recognition module that extracts features and assigns a label to each chip.

A recognition system is composed of an ordered set of feature detectors and a discriminant function to separate target
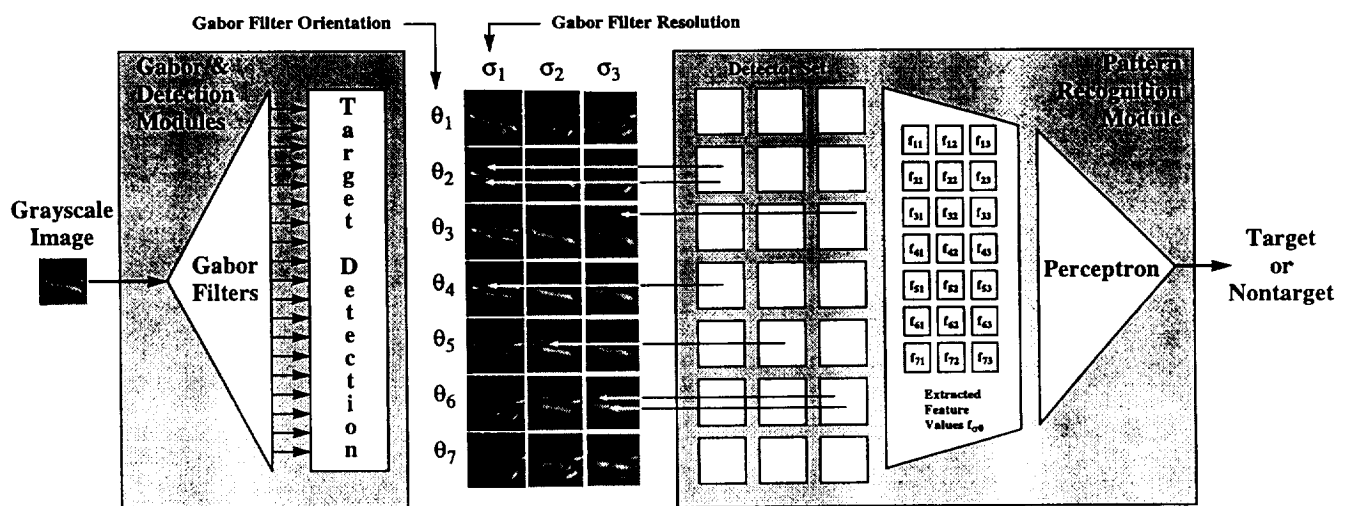

Figure 1. The representation of a pattern recognition system.

and nontarget stacks of chips. Each feature detector is represented by a convolution template large enough to cover the area of a chip. The convolution kernel contains a collection of probes points restricted to the values of +1 and -1. The use of a two-valued template allows detectors to explore both geometrical structure and contrast variation. For example, when a positive point (+1) is placed over a bright area of a chip and a negative point is simultaneously located over a dark area of the chip, the convolution operator produces a large output that signifies that the geometry and contrast variation embodied in the template exists at a specific displacement from the center of the chip. By adjusting the positions and values of the probe points, complex structural relationships can be readily identified. The set of detectors present in a single recognition system forms a registered set of convolution templates that serves as a 3D probe. The 3D probe spans multiple registered chips in the wavelet transformed stack and allows the detector set to explore relationships within a single stack-plane or across several stack-planes. For example, since the wavelet transform produces stack-planes corresponding to different spatial frequencies, the 3D probe can exploit multiple resolution levels to identify the presence of a fragmented edge that is not easily recognized at a high spatial frequency but is quite prominent at a lower spatial frequency. When the full set of detectors is convolved with a stack of images, an integer value feature vector is produced. Each detector contributes one value to the vector. By repeating this process for all the image stacks, a feature matrix is created that is passed to the discriminant function.

The E-MORPH system provides candidate detector sets to recognition system module and collects the corresponding system error as shown in Figure 2. The error vector is used to assess the accuracy of the candidate detector set and calculate a performance measure. Performance is defined using a combination of target and nontarget recognition accuracy (see Equation 1).

$$pm_i = \left(\alpha \cdot \frac{t_i}{T} + (1 - \alpha) \cdot \frac{nt_i}{NT} \right) \qquad \textbf{Equation 1}$$

The variable t_i is the number of targets correctly classified by the ith recognition system, T is the total number of targets, nt_i is the number of nontargets correctly classified, NT is the total number of nontargets, and α is a weight .

After all the recognition systems are assigned a performance measure, each recognition system competes for survival with other members of the population. The competition is organized as a tournament that ranks each recognition system based on its performance relative to the performance of other systems in the population. The size of the tournament changes throughout the evolutionary process and is based on the average performance of the population as shown in Equation 2.

$$NC = max\left(1, M \cdot \frac{\sum_{i=1}^{N} pm_i}{N} \right) \qquad \textbf{Equation 2}$$

In this equation, NC is the number of competitors in each tournament, N is the population size, and M is a user imposed upper limit on the number of competitions (M <= N). Each recognition system must win as many conflicts as possible to increase its chance for survival. The number of competitions won or lost is calculated using equation 3.

$$win_i = \sum_{k=1}^{NC} \left[\left(\frac{pm_i}{pm_i + pm_{2 \cdot N \cdot U(0,1)}} \right) < U(0,1) \right]$$ **Equation 3**

In these local competitions, the chance of winning is proportional to the ratio of the performance measures of the recognition system and its competitor. For example, if a recognition system's performance (pm_i) is high and a randomly selected competitor's performance ($pm_{2*N*U(0,1)}$) is low, then the probability that the ratio is greater than a value drawn from a uniformly distributed random variable U(0, 1) is also high. When the relation is satisfied, the recognition system wins the pairwise competition. Limiting the tournaments to a subset of the population reduces the possibility of premature convergence of the evolutionary process. When the average performance of the population is

Figure 2. Overview of the E-MORPH learning module.

poor, the number of individuals in each tournament is small and a marginally better recognition system does not have the opportunity to dominate the population. The pairwise competition used within each tournament tends to maintain a diverse population of recognition systems because marginal individuals always have a small probability of survival. The final selection for survival is based on a ranking of the number of conflicts won by each recognition system. The sets with the greatest number of victories survive to the next learning cycle.

E-MORPH uses two different techniques to alter the structure of the detector set contained in each recognition system. The position of the probe points in the convolution templates within a detector set are varied using evolutionary programming [Fogel, 1991], and the collection of convolution templates that forms a detector set is varied using a genetic algorithm [Holland, 1975]. These techniques are combined to explore different aspects of detector population. The evolutionary programming algorithm begins by cloning each member of the surviving population. The structure of the individual detectors within each clonal set is manipulated by vibrating the position of each probe point as shown in Figure 3.. The extent of the vibration of each probe point is modulated by a Gaussian envelope.

A complete E-MORPH learning cycle consists of several sub-cycles of the detector optimization using the evolutionary programming algorithm (EP) followed by several sub-cycles of detector set recombination using the genetic algorithm (GA) (see Figure 2). .The purpose of the EP algorithm is to systematically improve the position, type, and number of probe points in the active convolution templates. This is accomplished using a controlled vibration of the position of the points in each template followed by a series of random mutations that add and/or delete points (see Figure 3). To begin an EP cycle, the existing population of N detector sets is reproduced to form an extended population of 2N detector sets. Each detector set in the extended population is subjected to random variations. The amount of variation is inversely proportional to the performance of the parental detector set and controlled by Equation 4. The value $(x_{j,k}, y_{j,k})$ is the position of the kth point in the jth template, (X_{size}, Y_{size}) is size of the template, (X_{min}, Y_{min}) is the location of the lower left corner of the template, (X_{max}, Y_{max}) iS the location of the
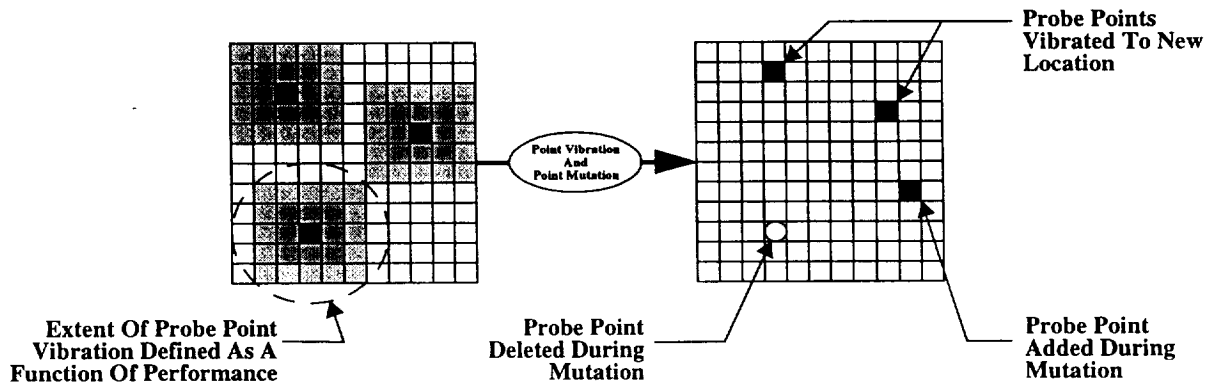

Figure 3. Using Evolutionary Programming to evolve a convolution kermel.

$$x_{j,k} = max\left(\left(min\left((x_{j,k} + \left(\frac{X_{size_j}}{2} \cdot 1 - pm_i\right) \cdot N(0,1)), X_{max_j}\right), X_{min_j}\right)\right)$$

$$y_{j,k} = max\left(\left(min\left((y_{j,k} + \left(\frac{Y_{size_j}}{2} \cdot 1 - pm_i\right) \cdot N(0,1)), Y_{max_j}\right), Y_{min_j}\right)\right) \qquad \textbf{Equation 4}$$

upper right corner of the template, $(1-pm_i)$ is the complement of the performance measure of the ith detector set, and $N(0,1)$ is a normally distributed random variable with a mean of zero and a variance of one. To update a probe point's position, the mean of the random variable is set to the value of the initial position of the probe point and the variance is scaled to fall into the range from zero to one-half the template size. Using this technique, when the performance measure is low, the potential extent of variation is high. The potential for variation is reduced as the performance increases. If the performance reaches one, the potential for variation is zero and the template's point configuration is frozen. This approach to adjusting the structure of a template is similar to the process of simulated annealing where gradual improvements in the population performance shutdown the process of random variation as a solution is formed.

The vibration process is only capable of adjusting the position of existing probe points. The second step of the EP phase is mutation that adds and/or deletes probe points to alter the complexity of the templates. Point mutation occurs immediately after the template points are vibrated. The amount of mutation is controlled by Equations 5 and 6.

$$p_a = max(\beta_a, \mu_a \cdot (1 - pm_i)^{\sigma_a}) \qquad \textbf{Equation 5}$$

$$p_d = max(\beta_d, \mu_d \cdot (1 - pm_i)^{\sigma_d}) \qquad \textbf{Equation 6}$$

The probability of a point mutation (p_a, p_d) is calculated using a user defined multiplier (μ_a, μ_d) that provides an upper limit on the mutation rate and a scale factor (σ_a, σ_d) that shapes the probability curve. In addition, the user must supply a baseline value (β_a, β_d) that sets a lower limit on the amount of mutation. For example, if the user sets $\beta_a=0.2$, $\mu_a=0.8$, and $\sigma_a=1.0$, then as the performance (pm_i) of the detector set rises, the mutation rate will start at 0.8 and fall off linearly to 0.2. By adjusting the value of σ_a, the rate of change of the mutation probability can be altered to remain high or low for larger ranges of performance. The mutation rate parameters must be set to correspond to the initial conditions used to form the population of detectors. If the detector set is initialized with a limited number of probe points, the probability of addition should be larger than the probability of deletion. This will bias the mutation rate toward addition and cause the detectors to grow in complexity.

After all of the templates in a detector set are vibrated and subjected to mutation, the set is placed in the recognition module and a Perceptron [Minsky and Papert, 1988] is trained to form a linear discriminant that separates target and nontarget chip-images. The error vector returned by the recognition module is used to assign a performance measure

to the modified detector set. This process is repeated for each detector set in the extended population. Finally, the N parental detector sets compete with the N offspring detector sets in a tournament for survival. The top ranked N detectors are preserved and the evolutionary programming cycle begins again. When the EP phase of E-MORPH terminates, each member of the population consists of a cooperative set of convolution templates.

The GA phase of E-MORPH exchanges convolution templates between pairs of detector sets to accelerate the learning process. To begin, a selection vector is formed to simulate a biased roulette wheel sampling process. The number of times a detector set appears in the selection vector is determined by the ratio of the set's performance measure to the average performance of the population. For example, if the the a detector set's performance is 0.8 and the average is 0.4, then the detector appears in the vector twice (0.8/0.4=2). If the ratio does not yield a whole number (e.g. 0.8/0.5=1+0.6), the remaining fraction is compared to a uniformly distributed random variable. If the fraction is larger, an additional copy of the detector is added to the selection vector. Performance is scaled prior to forming the selection vector to prevent a marginally better detector set from dominating the selection process [Goldberg, 1989] . Reproduction begins by drawing random pairs of parental detector sets from the selection vector. The set of convolution templates in each parent are analogous to a biological chromosome, and the individual templates are similar to genes. The uniform crossover operation consists of stepping through the detector sets of both parents, producing a copy of the templates at the corresponding positions, and assigning the copies to a pair of offspring detector sets at random. This is illustrated in Figure 4. The color coded (shades of gray) set of parental convolution templates is shown to the left and the mixture of templates duplicated in the offspring are shown to the right. The convolution templates are ordered to correspond to fixed template positions in the Gabored stack of chip images. When the crossover operation exchanges portions of the detector set, the potential exists to bring meaningful pieces
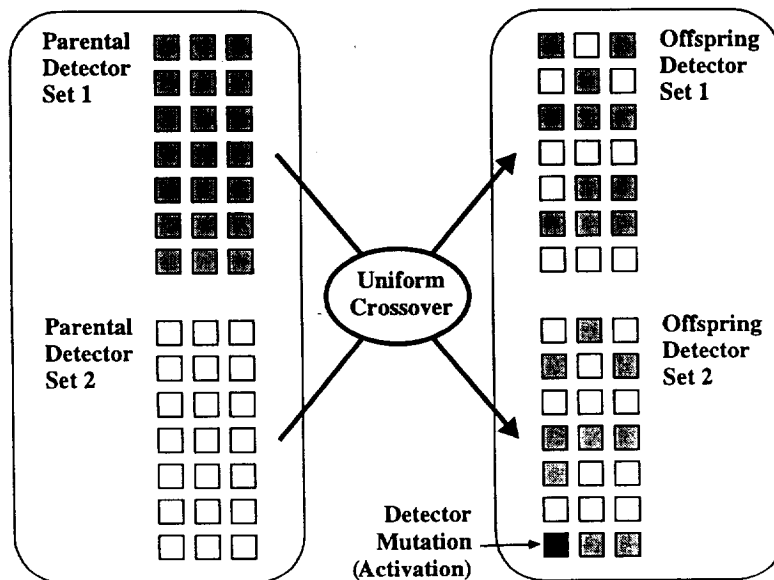

Figure 4. Detector set variation using Genetic Algorithms.

of two complex geometrical structures together.

After the recombination process, each offspring detector set is mutated by adding probe points to templates that are empty or by deleting all the probe points in a template. The EP phase is limited to vibrating and mutating convolution templates that are active (contain at least one probe point). The GA phase is responsible for activating and/or deactivating templates without altering the internal point distribution of active templates. The amount of addition and deletion is controlled by Equations 7 and 8 respectively.

$$p_A = max(\beta_A, \mu_A \cdot (1 - \overline{PM})^{(1 + (dt_i - \overline{DT})/\sigma_{DT})})$$ **Equation 7**

$$p_D = max(\beta_D, \mu_D \cdot \overline{PM}^{(1 + (dt_i - \overline{DT})/\sigma_{DT})})$$ **Equation 8**

These equations are similar to the distributions used to control point mutation during the the EP phase. A user supplied baseline mutation rate (β_A, β_A) insures that some level of variation continues throughout the learning process. The upper limit on the mutation rate (μ_A, μ_A) is used to avoid introducing an excessive amount of variation during the early stages of the evolutionary process before the performance begins to increase. The probability of activating or deactivating a template within a selected detector set is controlled by the average performance of the detector set population ($\overline{PM}$), the average number of activate templates ($\overline{DT}$) in the population, the variance (σ_{DT}) in number of templates in the population, and the number of active templates in the individual detector set (dt_i). Equations 7 and 8 scale the mutation rates to fall in the interval [β,μ]. The complexity term modulates the mutation rate by lowering it when an individual detector set' s number of activate templates exceeds the population average and raising it when the number is below average. These equations work to balance the opposing forces of performance and complexity by accelerating or decelerating the rate of variation of the detector sets.

The GA phase begins with N detector sets and combines N/2 pairs of parental set to form an extend population of 2N sets. Each member of the extended population is evaluated using the same procedure described for the EP phase. A tournament selection process is applied to rank the entire population and the N top-ranked detector sets are preserved for the next cycle of the GA algorithm. When the GA phase is complete, each detector set consists of combinations of templates that proved useful in the recognition process. In addition, the average number of active templates in detector set population will have evolved to produce detector sets with higher recognition accuracy.

The user can set parameters to control the number of EP and GA sub-cycles that occur within each E-MORPH learning cycle. There is a tradeoff between the EP and GA phases. The user can increase the sensitivity of the individual templates by increasing the number of passes through the EP phase relative to the number of passes through the GA phase. Alternatively, the user may elect to spend more computational resources adjusting the average complexity of the detector sets by increasing the number of passes through the GA phase. It is difficult to select an

appropriate mixture of passes because the evolutionary learning process is dynamic. During the early stages of evolution, it is not likely that the average number of active templates in the population is suitable for the recognition task. If the user arbitrarily increases the number of EP passes, the probe point density will increase to compensate for the lack of active templates. This will produce customized solutions that tend to perform well on training sets and poorly on test sets. If the number of GA cycles is too large, the average number of templates per detector will increase to compensate for the inadequate distribution of probe points within each template. Our solution to this problem is to use a large number of E-MORPH learning cycles and a small but equal number of passes through the EP and GA within each learning cycle. A better solution would be to implement an adaptive control mechanism that evaluates the relative contribution of each phase throughout the evolutionary process and dynamically adjusts the length of each phase.

EXPERIMENTAL DESIGN

To demonstrate our technique for generating a pattern recognition system, a target recognition task in medical imagery is presented. Specifically, the task is to locate and measure the deterioration in a patient's spinal column using radiographic images. This problem is difficult because the images contain large variations in contrast and the target vertebrae are extremely distorted. Examine the sample images shown in Figure 5. The vertebrae located in the
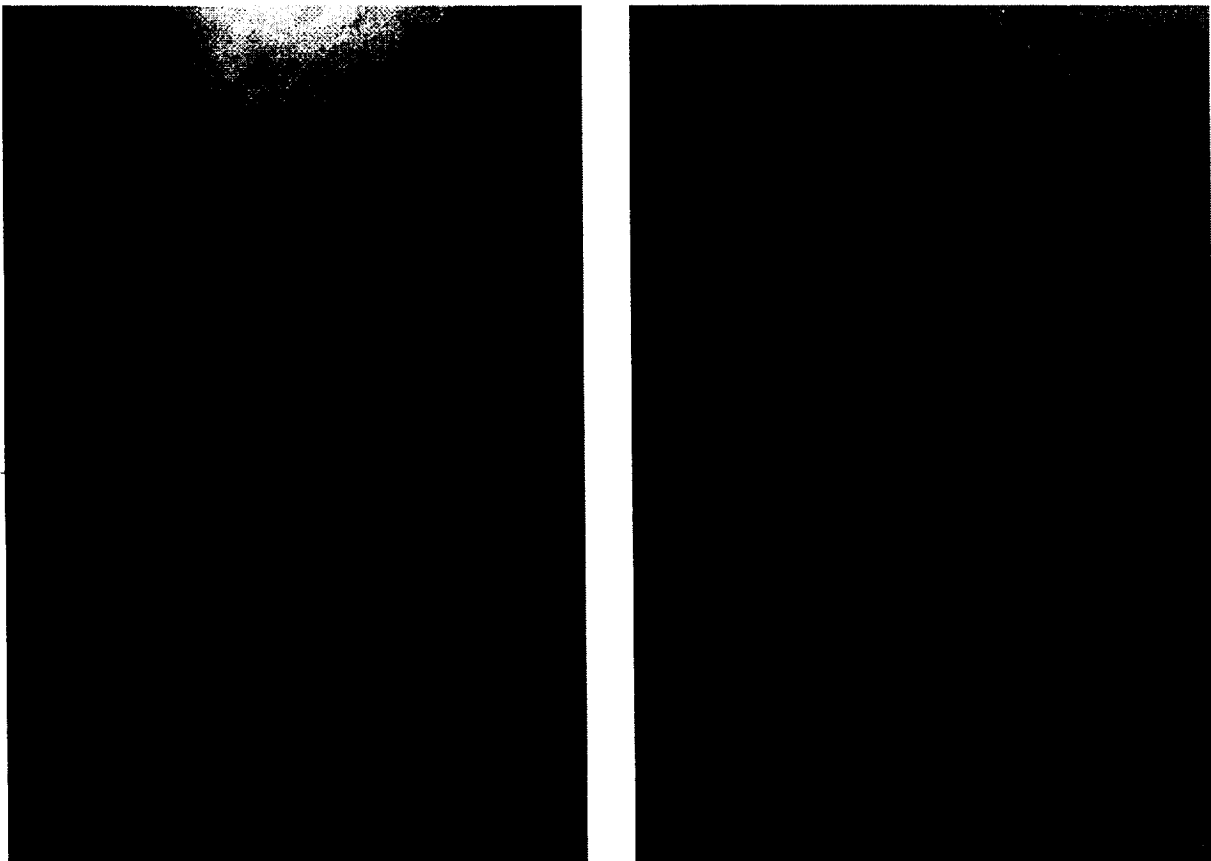

Figure 5. Sample x-ray images used in the target recognition experiment.

image on the right are slightly larger than the vertebrae in the image on the left. In addition, there are 3D effects caused by the tilting of the vertebrae as evidenced by the small elliptical shape visible on some vertebrae. To simplify this recognition task, we restricted our attention to the problem of locating the vertebrae. In particular, our goal is to detect the edges that define the gaps between vertebrae.

There were 48 variable size images used in this experiment. To begin, a central region measuring 512x768 pixels was extracted from each image. Each image was then processed using a set of two-dimensional self-similar Gabor functions. The Gabor functions are defined as the product of a radially symmetric Gaussian function and a complex sinusoidal wave as shown in Equation 9.

$$G(x, y; m, n) = e^{-r^2/S^2} \cdot e^{ikr} \qquad \qquad \textbf{Equation 9}$$

$$r = [x, y] \qquad k = \frac{1}{\lambda} \cdot \left[\cos\left(\frac{(3+m)\pi}{12}\right), \ \sin\left(\frac{(3+m)\pi}{12}\right) \right]$$

$$m = 0 \ldots 6 \qquad n = 1 \ldots 3 \qquad \lambda = 2 \cdot S = 4 \cdot n$$

The vector r represents a displacement vector and k is a propagation vector. The wavelength λ is specified in pixel units and is related to the width of the Gaussian envelope. The self-similar property of the Gabor function is evident in the resolution levels that are scaled versions of each other. In this experiment, we utilized seven orientations starting at 45° and continuing through 135° in increments of 15°. The resolution levels were set to λ equals 4, 8, and 12 pixels. The choice of parameters correspond to the general structure of the vertebrae in the grayscale images, but no attempt was made to optimize system performance by adjusting these values. Using these parameters, we generated 21 Gabor filters and applied them to the 512x768 grayscale images.

The Gabor output images are formed by convolving each Gabor filter with an input image. This is accomplished by computing the product of the Fourier transform of the Gabor filters with the Fourier transform of the image and taking the inverse Fourier transform of the product. This results in complex image consisting of real and imaginary values at each pixel location.

The real valued part of complex image corresponds to the cosine term of the propagation vector and the imaginary part is associated with the sine term. The cosine term produces a wave form that exhibits reflexive symmetry across the center of filter in the direction of the wave propagation while the sine term yields an asymmetric wave form. Both the sine and cosine act as edge detectors, but only the sine version is sensitive to the sign of the edge gradient. In this experiment, we use the magnitude of the complex image which produces a smeared version of the sine and cosine edge images due to the phase difference between these trigonometric functions. To complete the preprocessing, each of the resultant real valued magnitude images was byte scaled to map the dynamic range of the intensity into 256 discrete levels of gray.
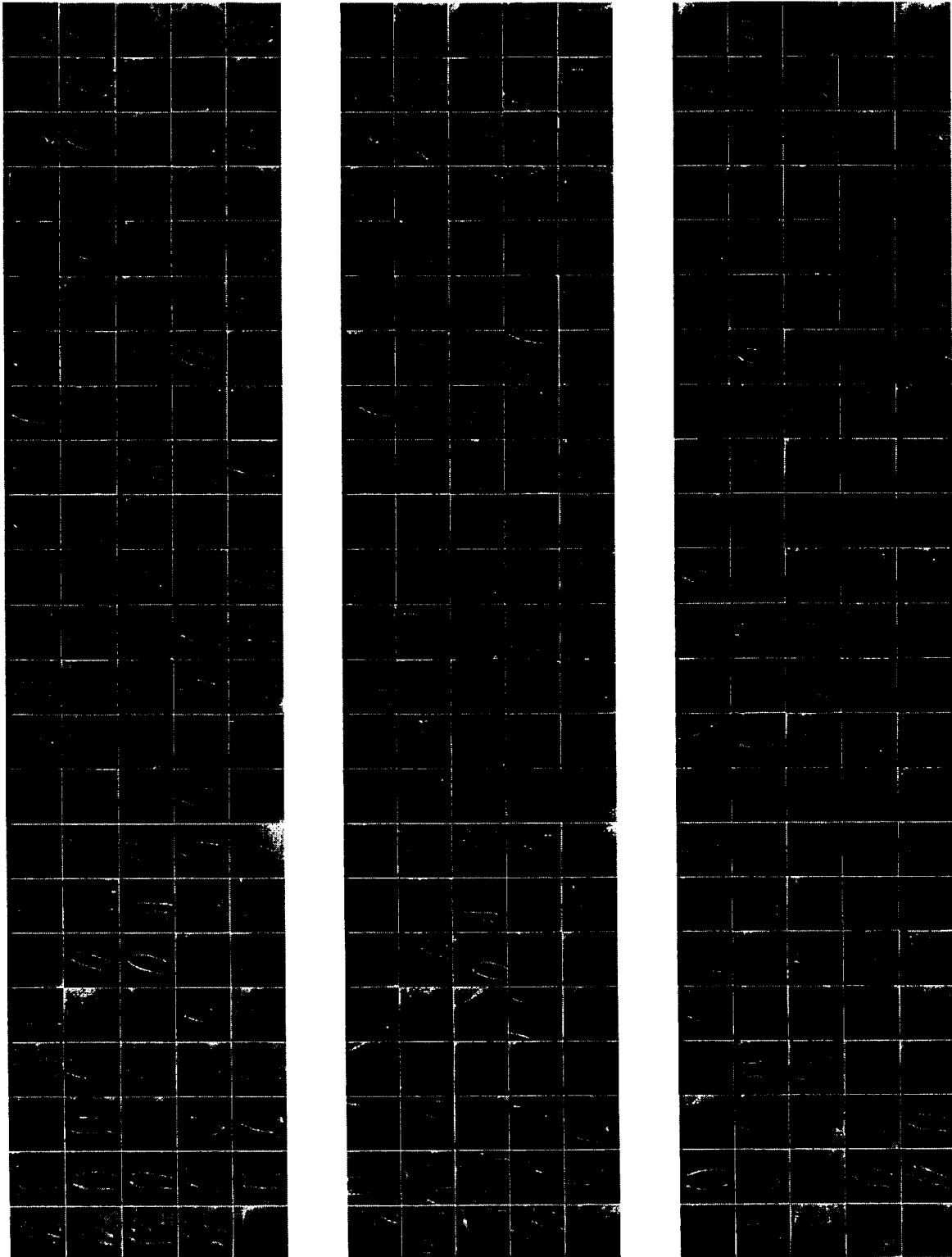
Figure 6. The training set composed of 345 grayscale chip images. The training set consists of 115 target chips (left column), 115 displaced nontarget chips (middle column), and 115 randomly selected nontarget chips (right column). Each chip is 128x128 pixels.
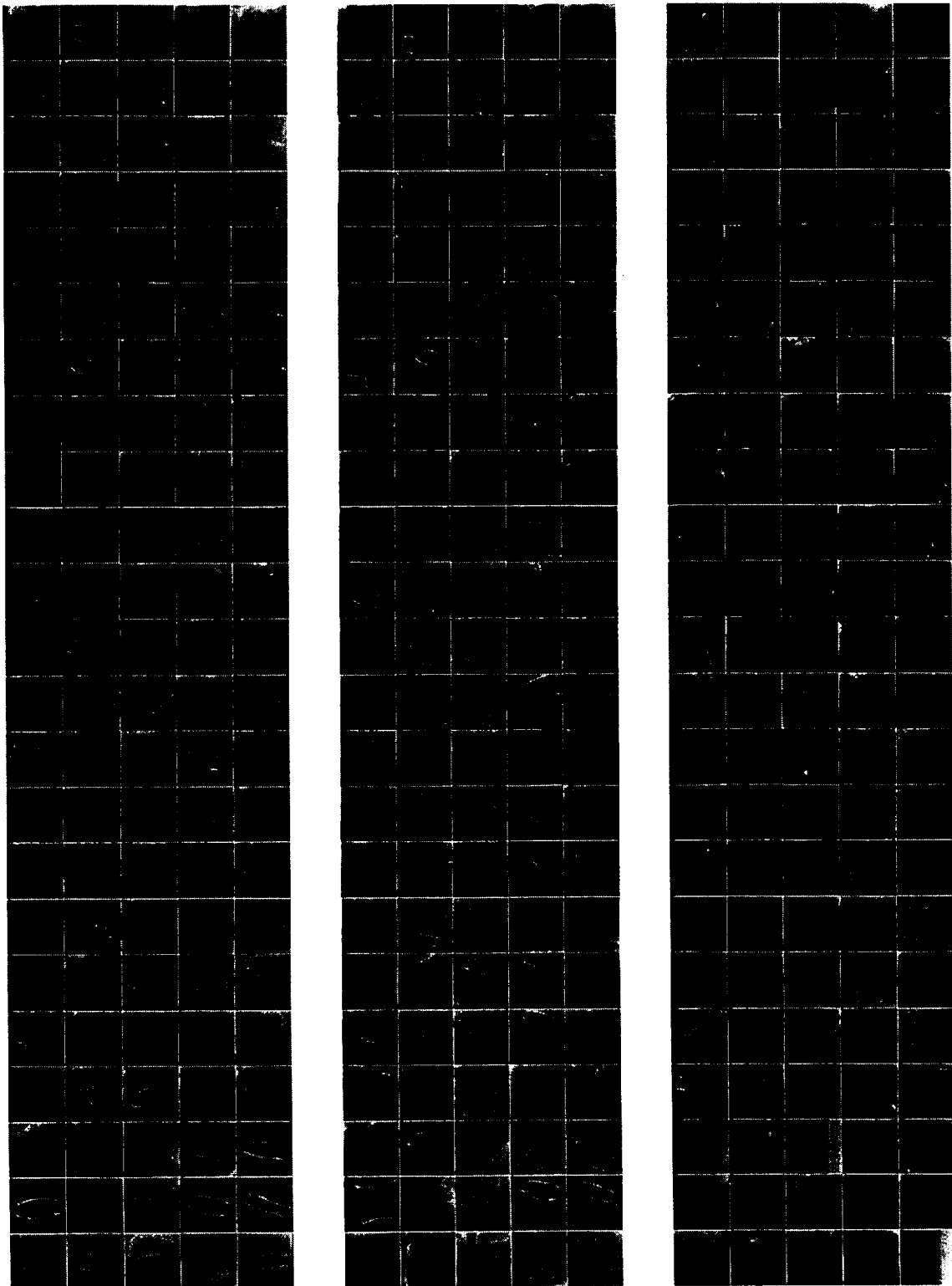
Figure 7. The test set composed of 345 grayscale chip images. The test set consists of 115 target chips (left column), 115 displaced nontarget chips (middle column), and 115 randomly selected nontarget chips (right column). Each chip is 128x128 pixels.
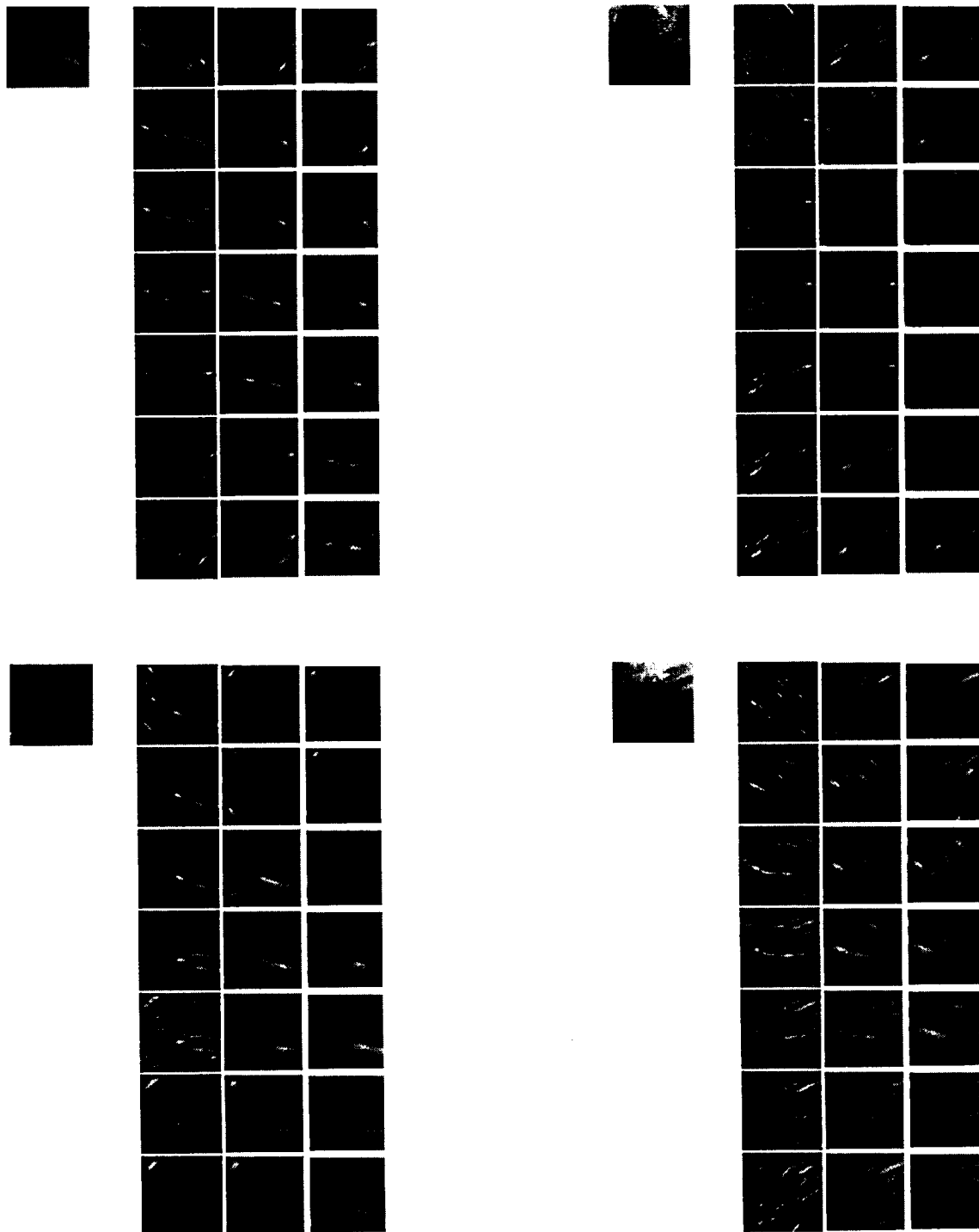
Figure 8. Sample Gabored chip-images. Four sample stacks of imgaes are shown. The small image to the left of each block of Gabor chips is the corresponding grayscale chip. The columns of eachblock of Gabor images represent different resolutions and the rows are different orientations (45-135 degrees in 15 degree increments). The top pair of blocks are targets, the bottom left block is a displaced nontarget corresponding to the target image directly above it, and the bottom right block is a random nontarget.

The automated target detection module is not implemented in this version of E-MORPH. Targets were located by examining individual raw grayscale image to identify the approximate center of each vertebrae gap. Chip images were then formed by defining a 128x128 bounding box around each target center. This process was repeated for each input image to produce 230 target chips. One set of nontarget chips was formed by displacing each target chip in some random direction by a minimum distance of 16 pixels and a maximum distance of 48 pixels. A second set of nontarget chips was formed by extracting 230 randomly located chips from the input images. When the three categories of chips (targets, displaced nontargets, and randomly selected nontargets) were combined, the result was a set of 690 images.

A training set consisting of 345 chips was formed by selecting 115 images at random from each of the three categories as shown in Figure 6. The remaining 345 chips were placed in a test set (see Figure 7). The grayscale version of the training and test sets are shown to aid the reader. E-MORPH does not operate on the raw grayscale images. The actual set of training images consisted of the 345 chip-stacks that were extracted from the corresponding positions in the Gabor filtered images. As a result, the actual number of individual chips in both the training and test sets was $21 * 345 = 7245$ chip images. A few sample Gabor stacks corresponding to the three categories of images (targets, displaced nontargets and random nontargets) are shown in Figure 8. Notice the amount of variation between the two sample target images shown at the top of the figure. Also observe the amount of similarity between a target image and its displaced nontarget counterpart shown at the left of the figure. The variation among the target images can easily exceed the variation between targets and displaced nontargets, making it extremely difficult to discriminate between target and nontarget images.

To begin the learning experiment, a population of 32 detector sets was generated. Each set was initialized by placing a probe point at a random location in three randomly selected templates. Theses templates were then evaluated by convolving them with the Gabor training chips to generate feature vectors that were passed to the Perceptron to compute a recognition accuracy for each detector set. The performance measure was computed by giving equal weight to target and nontarget accuracy .

An E-MORPH learning cycle consisted of five EP sub-cycles followed by five GA sub-cycles. The minimum and maximum values for the mutation rates in the EP phase were set to 0.2 and 0.6 for point addition and 0.05 and 0.2 for point deletion. The mutation rate limits for the GA phase were set to 0.2 and 0.6 for detector activation and 0.05 and 0.1 for detector deactivation. Each pass through a sub-cycle produced 32 mutated detector sets creating an extended population (parents and offspring) of 64 detector sets. The extended population was ranked using tournament selection and the top 32 detectors were saved to start the next sub-cycle. At the end of each pass through a sub-cycle, performance statistics for the population were saved. The experiment was run for a total of 22 learning cycles (110 EP sub-cycle and 110 GA sub-cycles).
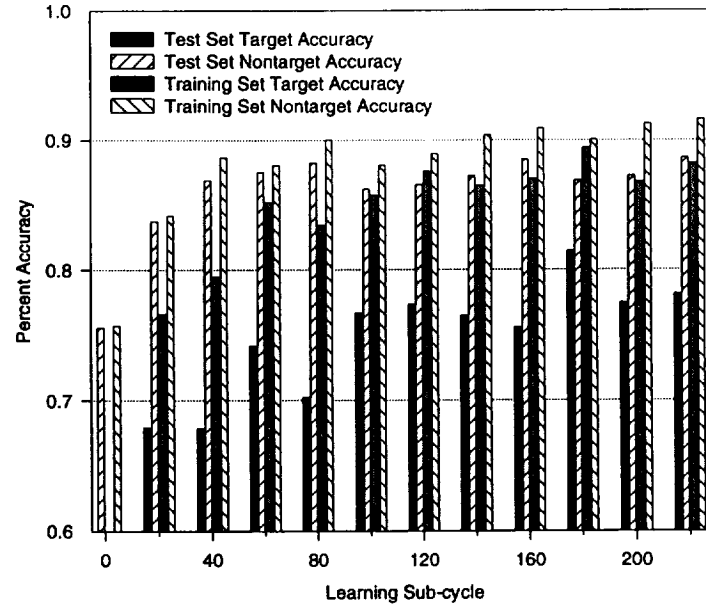
Figure 9. Average recognition accuracy.

The average recognition accuracy for the population produced during the evolutionary learning process is shown in Figure 9. The performance is displayed on even numbered learning cycles. Initially the average nontarget recognition was approximately 75%, but the target recognition accuracy was less than 10% (not shown on the graph). This high nontarget recognition accuracy is an artifact. Initially, most detectors produced a negative (< 0) response to almost every chip, and a negative response on a nontarget was recorded as a correct answer even though the detector had no ability to discriminate between targets and nontargets. After 20 sub-cycles, target recognition accuracy began to improve. Training accuracies were approximately 77% for targets and 84% for nontargets. The test set recognition accuracies lagged behind at 68% for targets and 76% for nontargets. As the evolutionary process continued, average recognition accuracies slowly increased to their f1nal values of 87% / 92% (target / nontarget) accuracy on the training sets and 78% / 88% (target / nontarget) accuracy on the test set. The difference between the training set and test set performance suggests that the detector sets are not generalizing to the test set. The difference between the training and test set target accuracy is the most obvious problem. The explanation of this difference is obvious if we review the response vectors for the individual detector sets at the end of the learning process (see Figure 10). The majority of errors occur among the target and displaced nontarget portions of the the training and test sets. To resolve these errors during the training process, the detector sets have to become highly customized. This specialization tends to limit the ability of the detectors to generalize when they are applied to the test set.

The best combined recognition accuracy achieved at the end of the experiment was 93% on the training set and 86% on the test set (see Figure 11). The fluctuations in the performance shown in this figure are due to the tournament selection process. The top performer in one generation can be eliminated i-rom the population during the competition

Targets   Nontargets I   Nontargets II   Targets   Nontargets I   Nontargets II
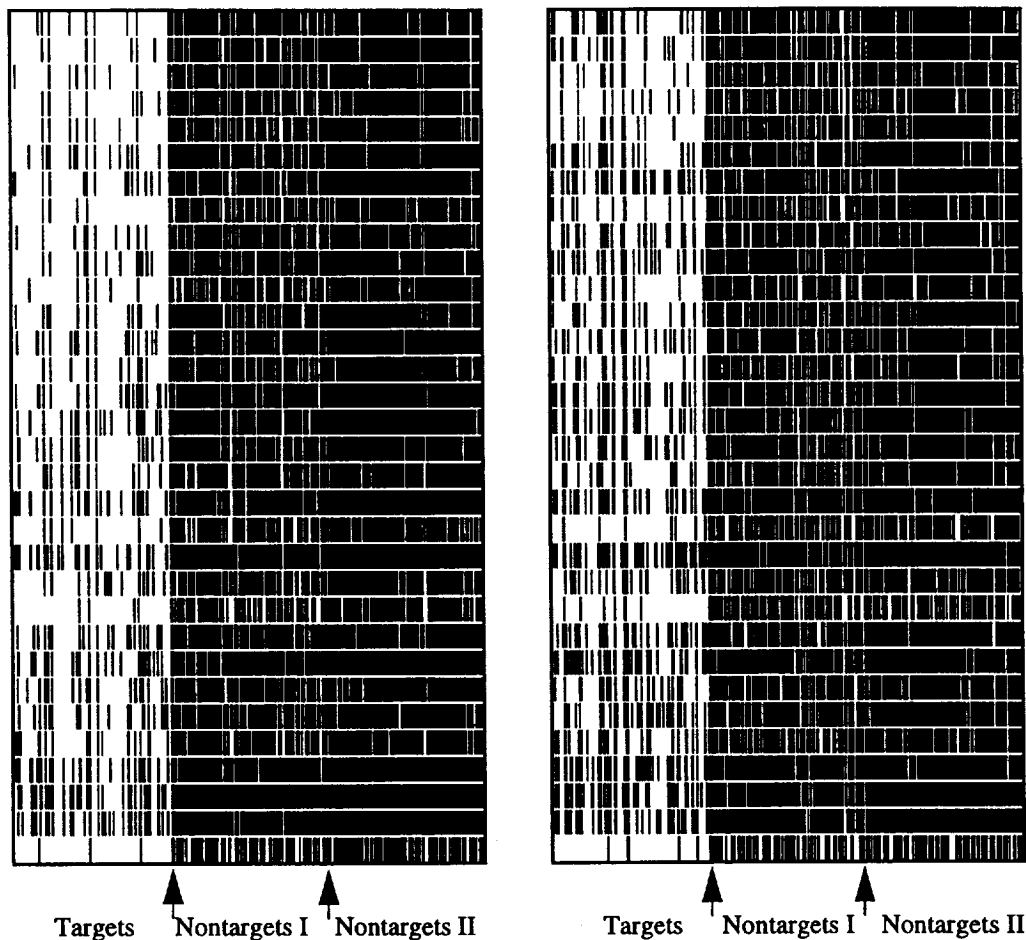
Figure 10. Final response vectors. The response vectors for the final population of 32 detector sets are shown. The training set response is shown in the left box and the corresponding test set response is shown in the right box. Responses are rank ordered by training performance. Each row represents the response of the detector set to 345 chip images grouped into targets, displaced nontargets (nontargets I), and randomly selected nontargets (nontargets II). A thin black stripe in the target area or a thin white stripe in the nontarget area represents one error.

for survival. Keeping the most accurate solution is one way to eliminate this problem, but notice the detector set that produces the best performance on the training set appears in sub-cycle 160 and has a rather marginal test set accuracy. Again this problem is related to the conflict between specialization and generalization caused by the similarity of the target and displaced nontarget images. One simple approach to deal with this problem is take a majority vote using the top three or five detectors. We tested this approach, and it increased recognition accuracy by 3-5%.

The structure of a few detector sets from the final population are shown in Figures 12 and 13. In Figure 12, three different detectors are superimposed on the first image in the training set. Notice the common substructure present in these three sets. This is to be expected at the end of the experiment as the system converges to a solution. What is surprising is the amount of subtle variation that still exists. For example, examine the rightmost template in the second to the last row of each detector set. These templates have the same basic footprint, but they are not identical. The basic pattern embodied in this template appears to align nicely with the vertebra gap, but it is difficult to predict

Best Recognition Accuracy
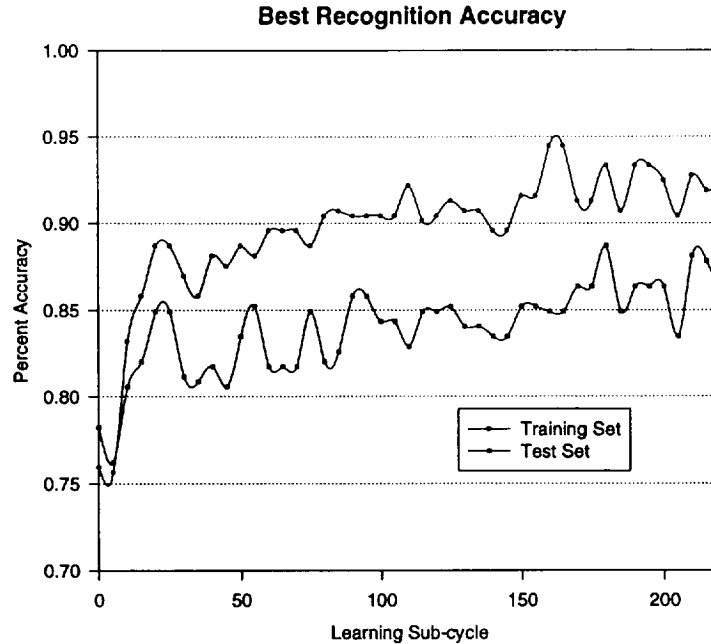
**Figure 11. The recognition accuracy of the best detector set.**

whether this is a useful template by examining a single Gabor image.

In Figure 13, one template is superimposed on two target images and a displaced nontarget image. This is the same detector set that appears in the upper left corner of Figure 12. Again examine the rightmost template in the next to the last row. Notice how the template's probe points align with the vertebrae edges in both targets but do not align with the displaced nontarget image. It is difficult to draw a general conclusion, but the distribution of points among different detector sets suggest that the cluster of points are sensing a on-off-on type of relationship embodied in the edge of a vertebrae. Occasionally, a cloud of points of the same type appear in a template as seen in the fifth row of the leftmost column. These clouds appear to sense large active or inactive areas in the chip images.

Although the points are scattered throughout the template. In general, a greater number of points appear in the lower resolution templates where edges are more pronounced. Also clouds of points are more common in the lower resolution templates. This suggests that the high frequency Gabor images are too sensitive to variations among the training images to be useful in the general solution, but they may be very useful in separating a target from its corresponding displaced nontarget image.

DISCUSSION

E-MORPH successfully generated a pattern recognition system capable of solving an extremely difficult problem in

Adaptive Vision Laboratory Final Report
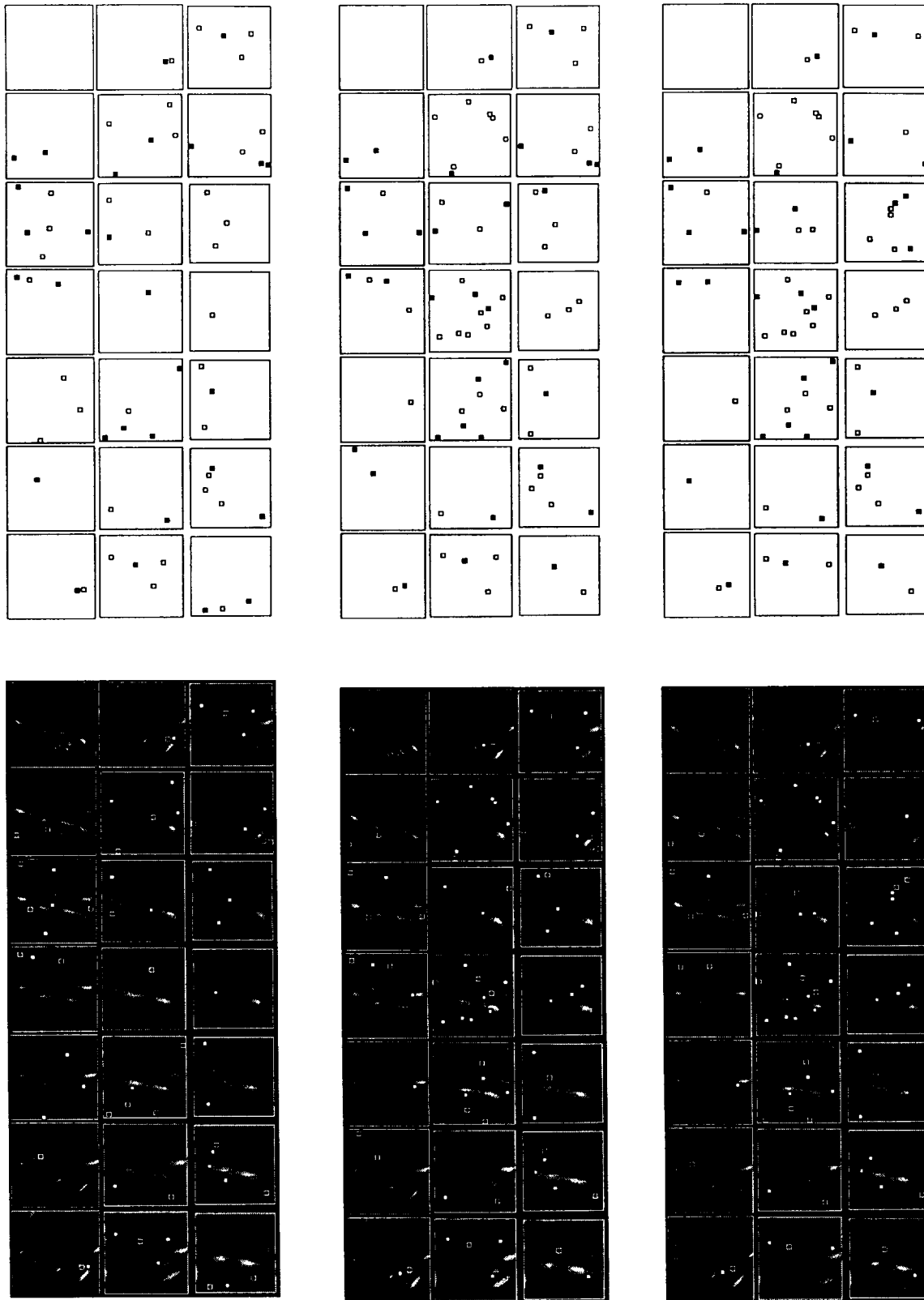
Figure 12 An E-EMORPH generate detector set applied to a target image. These three detector sets (top row) are taken from the population at the end of the final learning cycle. Each template probe point is shown superimposed on the corresponding Gabor target chip (bottom row). The dark center-white surround is a -1 point and the white center-dark surround is a +1 point.
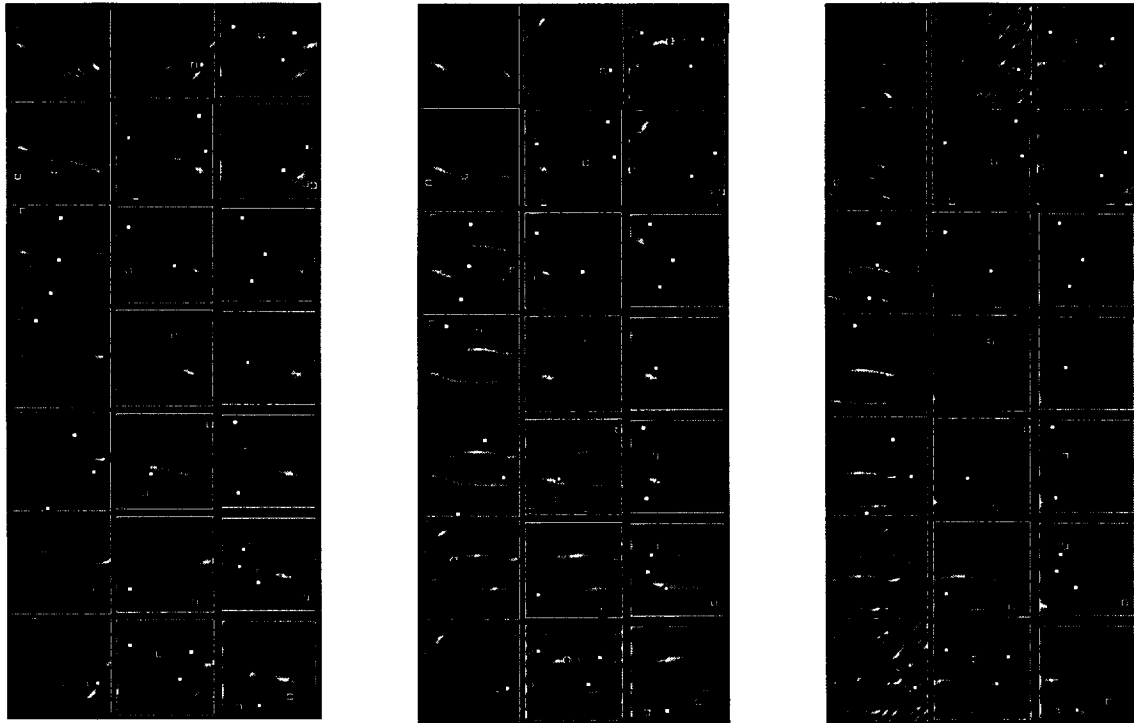
Figure 13 An E-MORPH generated detector set applied to three different Gabor images. A single detector set is shown applied to two target images (left and middle blocks) and a displaced nontarget image (right block).

medical imaging. In particular, x-ray images of human spinal columns are processed to locate vertebrae using Gabor filters to form a multi-resolution edge image. E-MORPH was then used to select features from the Gabor images and assemble pattern recognition systems. The learning process starts with a random assemblage of convolution templates that are enhanced using a hybrid evolutionary learning algorithm that exploits the strengths of both evolutionary programming and genetic algorithms. At the end of the experiment, the evolved population of feature detectors includes a detector that produces a 93% recognition accuracy on the training set and 86% accuracy on the independent test set. There are other members of the population that produce similar results. The performance can be improved by combining the results of several detector sets in a voting process, but a better solution is to add a more sophisticated classifier to deal with the in-class variability of the data set. Clearly, the capability of the linear discriminant used to separate targets and nontargets is limited and forces E-MORPH to compensate by generating detector sets customized to the training data.

The overall structure of the detectors generated by E-MORPH appears to correspond to both the geometrical and contrast variations present in the images, but the complexity of the training set makes it difficult to analyze the behavior of the individual detector sets. We believe the detector sets are using a complex combination of geometric structure, contrast variation, and statistical averaging of the lower spatial frequencies present in specific Gabored images to guide the search process.

There is no single approach that solves all problems in automatic target recognition. E-MORPH represents one viable alternative. Solutions generated using our evolutionary learning algorithm are quite different than solutions produced by human experts. This suggests that human experts may not be using all of the available information to develop robust pattern recognition systems. In future work, we hope to explore the possibility of combining human expertise with the evolutionary search process to access these design alternatives. This hybrid approach to design may ultimately produce recognition systems with performance superior to any in use today.

REFERENCES

Fogel, D. B . (1991). *System Identification Through Simulated Evolution: A Machine Learning Approach to Modeling*, Needham, MA: Ginn Press.

Gabor, D. (1946). "Theory of Communication", J.I.E.E., 93:429-459.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*, Reading, MA: Addison-Wesley.

Holland, J. H. (1975). *Adaptation in Natural andArtificial Systems*, Ann Arbor, MI: The University of Michigan Press.

Minsky, M. L. and S. A. Papert (1988). *Perceptrons*, Cambridge, MA: The MIT Press.

Rizki, M. M., L. A. Tamburino, and M. A. Zmuda (1993) Evolving multi-resolution feature detectors. In *Proceedings of the Second Annual Conference on Evolutionary Learning*, eds. D.B. Fogel and W. Atmar, La Jolla, CA: Evolutionary Programming Society, 57-66.

.Rizki, M. M., L. A. Tamburino, and M. A. Zmuda (1994) E-MORPH: A two-phased learning system for evolving morphological classification systems. In *Proceedings of the Third Annual Conference on Evolutionary Learning*, eds. A. V. Sebald and L. J. Fogel, River Edge, NJ: World Scientific, 60-67.