AUTOMATED FEATURE DETECTION
USING EVOLUTIONARY LEARNING PROCESSES

Louis A. Tamburino
Wright-Patterson AFB

Mateen M. Rizki
Wright State University

William Van Valkenburgh
Wright State University

ABSTRACT

The design and implementation of a software environment for the study of learning strategies applied to tasks in image processing is described. This environment facilitates the systematic exploration of evolutionary learning processes and embedded adaptive control mechanisms that modify the process of feature extraction. Selected principles of mathematical morphology define the nucleus of this image processing-learning system. A preliminary experiment with a two-class recognition system is described. Initial observations are discussed for a recognition task that requires the classification of uppercase English letters into two categories: target or nontarget.

INTRODUCTION

A classical pattern recognition system is composed of a set of feature detectors and a classifier. The feature detectors respond to the presence of specific patterns in images and these responses are fed to the classifier that assigns each image to a category. Often, a detector responds to a local feature that occurs in several images. Using the responses from a set of detectors, the classifier partitions feature space into a set of regions that correspond to a desired set of image categories.

Feature detectors are problem specific. Sets of detectors are tuned to specific images and objects in those images. For example, detectors that extract useful information from images of aircraft are probably not good detectors for extracting information from images of tissue cultures. Detecting edges is useful in the first case while sensing texture is probably more important in the second case. Clearly, different features are important in different problem domains and normally a human expert is required to identify these crucial features. There is an alternative solution -- construct a learning system that automates the process of feature selection and produces a feature detector to perform the required classification.

In this paper, a simplified recognition system is presented for the classification of binary images into two categories: target and non-target. This recognition task is accomplished by constructing a single feature detector that only responds to images in a selected target class. When the detector is applied to images that are not members of the target class, there is no response. Consequently, a classifier system is not necessary in a two-class recognition problem because the binary response of a single feature detector identifies the class of each image.

AUTOMATED FEATURE DETECTION

Automating the process of feature detector construction requires a feedback system that applies feature detectors to images, evaluates their response, and adapts the detectors to increase their level of discrimination. We present a learning-recognition system composed of two components: a simplified recognition system and a restricted evolutionary learning system. The recognition system contains a single feature detector. This detector is applied to a set of images and the response is recorded. The evolutionary learning system evaluates this response and generates a modification to the detector. This process repeats until a predetermined limit on computational resources is reached. An independent test set of images is then used to evaluate the final quality of the feature detector.

The form of feature detectors used in this system is based on the principles of mathematical morphology (Serra, 1982). In our prototype system, morphological analysis is limited to erosion operations that are specialized probes used to evaluate binary images. In this process an input image (I) is eroded by a set of probes (S) to produce a resultant image (Eq(1)).

Erosion (special case) E: I x S -> I Eq(1)

Specifically, erosion is the systematic removal of points from an input image as it is compared to a
A structuring element is a collection of points that serves as a template that is translated across the surface of the input image. One point in the structuring element is designated as a reference point. Each coincidence between the template and the image is marked in a resultant image, at the position of the reference point. The result of eroding an image by a structuring element is the production of a resultant image containing a smaller collection of points called the footprint. An example of the operation of erosion is shown in figure 1. Here three different images are eroded by the same structuring element. The black points define the structuring element and the point marked with an X is the reference point. The structuring element fits in the letters L and E but does not fit in the letter F. When a fit occurs it is indicated by a non-null response (R=1). The three responses combined form a feature response vector (110). Note, the resultant images contain only the points marked with the letter X. The full structuring element is superimposed on the images to aid the reader in understanding where the structuring element fits in the image. Erosion is implemented as a series of logical SHIFT and AND operations Eq(2). Let $S \in \{0,1\}^{(k\times k)}$ represent a structuring element and $\delta$ be a displacement specified by a point in the structuring element.

$$E(I, S) \Rightarrow I = \cap I_{\delta}$$

where $I_{\delta} = \{x; x - \delta \in S\}$ specifies a shift operation.

Each point in the structuring element defines a displacement relative to the reference point. The order of evaluation of the displacements is irrelevant and the result is shift invariant.

The erosion operation uncovers only patterns contained in images. Frequently, it is necessary to decide that a pattern is not contained in an image. For this process, a hit-or-miss template is defined that is composed of two types of points: foreground (fg) and background (bg). Foreground points are processed as described in Eq(2), whereas background points are required to fit in the background of the input image (Eq(3)). The background fit is calculated by eroding the complement of the image (I) using the points in the background portion of the structuring element as displacements. The introduction of background points facilitates the detection of edges and holes in images. In figure 2, a single background point is introduced into the same structuring element that appeared in figure 1. This point is marked by a white box superimposed on a black cell. When this modified structuring element is applied to the same three images, the background point eliminates the upper bar of the letter L producing a response vector of (100). If our images are restricted to the letters L, E, and F, then the structuring element displayed serves as an L detector.

$$H(I, S_{(fg,bg)}) = E(I, S_{fg}) \cap E(I, S_{bg})$$

Eq(3)

The learning system navigates the space of potential structuring elements using feedback information provided by an evaluation system. In the experimental work described in this paper, the problem is to construct a 16 x 16 pixel, binary feature detector capable of discriminating between a target and non-target class of 16 x 16 pixel, binary images. There are approximately $2^{256}$ possible structuring elements containing only foreground points. The learning system must successfully locate a single structuring element in this search space using a global constraint: the selected structuring element must fit into all the target images while minimizing erroneous responses.

An evolutionary learning strategy is used to navigate the space. This process is analogous to biological evolution where successful individuals reproduce with variation and competition within the population selects those individuals that survive to reproduce. In this metaphor, structuring elements are organisms that reproduce with variation. The process of reproduction with variation is modeled as the extension of successful structuring elements by the addition of new points. A fixed sized population of the best structuring elements is retained and new structuring elements compete to enter this population.

To simplify the preliminary experiment, a restricted evolutionary learning process is implemented. This process consists of generating random points to add to a structuring element. As each point is generated, a local evaluation is used to identify non-viable points. After a predetermined number of points are tested, the generated structuring element is compared to the population of top performers. If the new structuring element surpasses the population's minimum performer it will replace it. This restricted evolutionary search process is equivalent to a guided random search strategy.

TWO-CLASS RECOGNITION EXPERIMENT

This experiment begins to explore the capabilities of the learning-recognition system as it applies to a simple problem: design a two-class recognition system that discriminates a target uppercase letter from other letters in the English alphabet. A number
Figure 1. Erosion Using Foreground Points. The erosion of several images by a foreground structuring element. The reference point of the element is represented by an X'ed black pixel. The resultant image has a pixel remaining wherever the reference point is located when the element fits within the initial image. The rightmost images show this by overlaying the element upon the initial images. The binary response (R) values at the far right are one only if the element fits within the image.

Figure 2. Erosion Using Foreground and Background Points. The erosion of several images by a hit-or-miss structuring element. Background pixels are represented by a white box superimposed on a black cell. The resultant image of an erosion by a hit-or-miss template has a pixel remaining wherever the foreground fits within the image and the background does not. This is shown in the rightmost images by overlaying the structuring element upon the initial images. The binary response values at the far right are one only if the structuring element fits within the image.
of researchers have recently investigated this problem using a variety of techniques (Stentiford, 1985; Fukushima et al, 1983). Unlike these researchers who are addressing practical applications, our approach is restricted by our limited computational resources to the simpler problem of constructing a single binary feature detector capable of exclusive recognition of a target. We selected this problem because it is sufficiently complex to serve as a test bed for studying fundamental issues in learning, yet, simulation runs require only a few hours of computation on a microcomputer (Zenith 248).

Each training experiment operates on an image set of 128 characters (figure 3). These characters are divided into 16 subsets, each containing 8 examples. Some examples are slightly distorted and/or contain random noise. Two parameters control the length of an experiment: the amount of resources available to generate a single structuring element, and the total resources available to generate all structuring elements. The basic unit of resource is a binary word operation. One word operation corresponds to the cost of forming the logical AND of two words. When an image is eroded by a structuring element, the action of ANDing collections of bits in images consumes a definite number of word operations. The relationship among total resources, resources per structuring element, and system performance are discussed in Tamburino, et. al (1989).

We begin with an empty structuring element and a population of size zero. Candidate structuring elements are generated using the guided random search algorithm described above. Points are generated, evaluated, and accepted or rejected. The process of evaluation consists of eroding each image in the training set by the structuring element. A new performance is calculated as the sum of the number of responses to the target class images multiplied by the number of nontarget classes less the number of responses to nontarget images. If the performance is improved, the point is added to the structuring element. As point are accepted, errors in the response vectors are eliminated and the sensitivity of the detector is increased. The cost of evaluation is the same for viable and non-viable points. Each point requires the application of the erosion operator, and this process consumes word operations. Consequently, both successful and unsuccessful points reduce the resources allotted. When all the resources earmarked for a single structuring element are spent, the final performance achieved by the new structuring element is compared to the population of top performers. If the new structuring element's performance is high enough, it replaces the population element with the lowest performance. This entire process is then repeated until the total resources for the experiment are expended.

At the end of each experiment, the top performing detectors are applied to an independent test set (figure 4). This set has the same composition as the training set: 128 characters divided into 16 subsets. The response of the structuring elements to this set determines how well the solution may be generalized.

Figure 3. Sample Training Set.
PRELIMINARY RESULTS

Results are presented from four training sessions. In the first session, the target image set is composed of eight samples of the letter E and 120 other letters or non-target images (see figure 3). Five million word operations are allocated to this run with a limit of 8000 word operations per structuring element. The structuring element is restricted to foreground points. In figure 5a, the training and test performance of the best 33 structuring elements is displayed. The structuring elements are ranked based solely on their training set performance. Observe that a structuring element with perfect performance on the training is not necessarily the best performer on the test set. There appears to be a tradeoff between feature detector generalization and specialization. As a detector is tuned to the training set, the idiosyncrasies of the specific training images are incorporated into the detector. Consequently, the detectors that are most discriminating on the test set are slightly below peak performance on the training set.

In a second training session, the same experiment is repeated but the word operations allocated to each structuring element are increased to 22,000 per structuring element. Since the total word operations allotted to each experiment is constant, this permits the creation of a maximum of 227 structuring elements as opposed to the first experiment where 625 structuring element are constructed. A marginal improvement in training performance is noted but the test performance is significantly lower (figure 5b). The increased word operations per structuring element are spent fine tuning fewer structuring elements. Consequently, the structuring elements are customized to the training set and test set performance is reduced. The drop in test performance may be linked to the reduction in the number of structuring elements explored. Since structuring element construction is strictly additive, the first few points placed in the element strongly influence the final result. This effect becomes noticeable as the number of structuring elements tried is reduced.

Figure 6 shows the first eighteen structuring elements constructed in session 2. Contrast the first structuring element and the tenth structuring element. Notice the tenth element recognizes longer horizontal top and bottom bars than the first structuring element. This allows the tenth element to distinguish a larger variety of the letter E from the letters B and D. The short top bar that appears in the first structuring element is customized to the training set.

In the third session, the target image set is the letter Y. Again, five million word operations are allotted to the run with a limit of 8000 word operations per structuring element. The structuring element is restricted to foreground points. In figure 7a, the training and test performance of the best 33 structuring elements is displayed. The training performance is significantly lower than the results obtained for the letter E. Certain letters are more
difficult to recognize. Y is the most difficult in the training set; the point of intersection of the three segments of the letter Y varies widely in different images. This variation creates interference between the letter Y and the letters N, X, and Z.

In the final session, Y is again selected as the target image but the structuring element is constructed with both foreground and background points. The resource allocation is the same as that used in session 3. In this case, both training and test performance are raised significantly (figure 7b). With the addition of background points, the system is able to mask portions of the interfering letters. Consequently, the distorted cases of the letter X are eliminated by a few well placed background points that restrict the diagonals of the letter Y.
SUMMARY

Although these are not definitive experiments, a number of interesting observations are noted.

(1) Different allocations of total resources and resources per structuring element significantly alter the results. This implies that the level of difficulty of this problem is sufficient to merit the application of learning strategies.

(2) The distribution of resources in the learning system is critical: too little or too much produces flawed results. Giving the system excess resources produces structuring elements that are overtrained, while giving the system too few resources, produces structuring elements that cannot perform the required task. The tradeoff between specialization and generalization is a fundamental issue for future research.

(3) Finding a single global feature detector is highly restrictive. If there are large variation in the image set, a single detector is prone to error. Using the techniques described, we can create additional independent detectors and combine their response vectors using conventional logic operations. In effect, we poll several detectors to increase the accuracy of the response.

(4) The training set performance is not a good indicator of test set performance in this problem. This is possibly due to a failure in the experiment setup. Overall, the characters in the training set (figure 3) are thicker than the characters in the test set (figure 4). This means that the structuring elements are not only recognizing characters, but also, the thickness of characters. In future experiments, this bias will be eliminated by creating training sets using a random selection of characters.

(5) The selection criterion for point acceptance is based on an increase in performance. We have explored a less restrictive criterion that accepts any point that does not decrease performance. We found that this alternative selection criterion increases the point density of the structuring elements while maintaining a comparable level of performance on the training set. Since our goal is to capture invariant properties of the uppercase letters, we desire the less dense structuring elements produced by the restricted performance measure. To expand the learning system, many new selection and evaluation criteria are required, and these criteria must compensate for bias in the image sets.

The primary focus of our research is the study of learning strategies and evaluation techniques. We use image processing applications to provide a sufficiently complex study environment. This work establishes the operating principles of an evolutionary learning process as applied to a restricted problem in
image processing, while testing the operation of a specialized software environment. It serves as a starting point for further experimentation.

ACKNOWLEDGEMENT

We would like to thank Dr Robert Trenary for his early contributions to this project. Special thanks to Michael Zmuda for his outstanding software support.

This work is supported by the Air Force Office of Scientific Research, Directorate of Electronic and Material Sciences under task 2305R5 and contracts F33615-87-C-1416 and F49620-88-C-0053.

REFERENCES


