A Staged Genetic Programming Strategy for Image Analysis

Daniel Howard and Simon C. Roberts
Software Evolution Centre, dhoward@dera.gov.uk, Tel:(44)(1684)894480,
Building U50, Systems & Software Engineering Centre,
Defence Evaluation and Research Agency,
Malvern, WORCS WR14 3PS, UK.

Abstract

The problem addressed is that of automating the task of visual inspection of images to detect objects of interest. Detectors are functions of the pixel information. When moved across an image these discriminate objects from non-objects according to the value that is computed. The detector is evolved off-line in two evolution stages. This strategy results in practical evolution times. It produces fast automatic detectors that can be interpreted to understand object detection principles.

1 Introduction

An attraction of Genetic Programming is its ability to return a tangible formula or computer program that can shed light into the logic employed to solve a task. For the problem of machine vision, an analysis of this logic may explain some general principle for discriminating an object from a scene.

The method discussed in this paper uses a number of examples of 'object' and 'non-object' to evolve a discriminator or object detector. The evolution is an inductive learning and CPU intensive task that is carried out off-line to discover a detector that can be applied to a great number of images on-line.

A potential problem area with GP is the growth of its structures, however, this appears to relate to its power as a search method. This growth can significantly augment the computing requirements of evolution. It also means that large detectors can be difficult to interpret even after they have been simplified algebraically; detectors will contain more operations; and possibly take longer to process an image.

A staged evolutionary process is developed in section 2 to circumvent some of these difficulties. The validity of this approach is discussed in section 3 with reference to a SAR image analysis problem. An IR image analysis task in section 4 illustrates how the approach can begin to address the so called stability-plasticity dilemma [1].

2 Staged approach

Given a rectangular image of dimensions N x M pixels, containing objects that can be bounded in subwindows of size n x n pixels, the task is to construct a function of pixel data with support n x n that can be systematically moved across the entire image, i.e. is evaluated pixel by pixel and line by line. Typically, N x M is $O(10^6)$ pixels and n x n is $O(10^2)$ pixels.

At each pixel the function returns a value that is either positive (P) or negative (N) to either denote detection of an object or of a non-object respectively. Usually, the number of non-object pixels dominates, i.e. the number of object pixels in N x M is very small.

A number of such images is selected to evolve the detector, and a 'truth' of known object locations is prepared for each image. A true positive (TP) results when the detector correctly detects an object from the truth; a false positive (FP) occurs when it incorrectly detects an object - also known as a 'false alarm'. Similarly a true negative (TN) is a correct non-object detection and a false negative (FN) is an object missed.

When an object is fairly small, e.g. n x $n = O(10)$ as in section 3, it is appropriate to denote the object as a single pixel. For larger objects, e.g. as in section 4, a small area of pixels or a line of pixels may be used.

The evolution process that attempts to discriminate every object pixel from every non-object pixel in the image is prohibitively expensive as it requires repeated function evaluations at all N x M pixels.

In the staged GP method presented here, the first stage of GP attempts to discriminate every object pixel from a small random selection of non-object pixels. Upon completion of the first stage of GP, the fittest detector is applied to the entire image, i.e. all N x M pixels. This results in a number of misclassifications or FP - non-object pixels that are similar to object pixels. In a second stage, a brand new run of GP now attempts to discriminate every object pixel from these discovered FP.

The fittest detector from the first stage is combined or 'fused' with the fittest detector from the second stage. It is done such that both detectors must detect the object for it to be counted as detected, i.e. both must return a positive value. When applying them to an image on-line, the first detector is evaluated first at all N x M pixels, and the second detector is applied only to the positive returns from the first - a very small number of pixels relative to N x M. It is important to note that the second evolution stage has a tougher job than the first stage because it must discriminate like from like. Another observation is that there is little point in varying the random set of non-object pixels used by the first stage because 'difficult' non-object pixels are few and unlikely to be covered with such a small random set of non-object pixels.

It turns out that one can exploit this situation to circumvent the program growth and some of the apparent disadvantages of GP. Because the first detector so dominates the CPU time of on-line image processing it is essential that it be a small and efficient function. This is achieved by setting a low maximum size of tree and limiting the variety of terminal nodes in the first evolution stage. These constraints do not compromise the power of the overall evolution process because only the second evolution stage has the difficult job of discriminating like from like, and it has the freedom to evolve large trees with many types of terminals to achieve its task.

As a result: an efficient on-line object detector is produced and also a first detector is produced that can always be interpreted as it is very small, e.g. 20 nodes. Evolution times are also reasonable because the staged evolution process involves a minority of pixels from the N x M set.

The two stage GP process can be generalised into a multi-staged process to address additional requirements as discussed in section 4.

3 Two-stage GP

The experiments in this section demonstrate the utility of an evolution strategy involving two GP stages in a real world image analysis problem setting. The reader can consult references [2-4] for a more complete account than is provided in this section.

3.1 Test problem and GP formulation

The objective was to evolve an automatic ship detector for Synthetic Aperture Radar (SAR) images of the English Channel taken by the ERS-1 satellite both at 50 metre and 100 metre resolution. In order to produce a detector that generalised well, the set of images was divided into three: training, test, and blind sets. These consisted of one, two, and two images respectively. Figure 1 shows two of these images that are typical. Each image is made up of approximately one million grey level pixels.

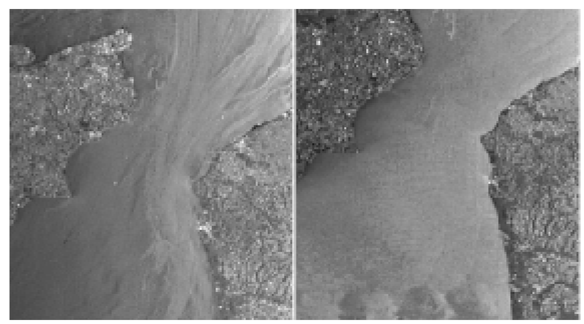

Figure 1: SAR images of the English Channel.

The objects in this problem are essentially very small and were represented in the truth at single pixel locations. The n x n subwindow as defined in section 2 was kept small, i.e. it never exceeded 9 x 9. Figure 2 shows some of these ships in-situ. A steady-state GP implementation, was run with MS Visual C++ under Windows NT and on a number of Pentium II PCs. It was modified to work with pre-computed statistics of pixel data as the GP primitives - a technique discussed by Poli [5].

Figure 3 illustrates some of the GP primitives defined in the n x n sub-window. For example, $I9$ is the average of the pixel values in a 9 by 9 box centred on the pixel; V9 is the standard deviation of the pixel values in a 1 pixel perimeter on the edge of this 9 by 9 box; $P9$ is the corresponding average of the same area; while $D37 = I3 - P7$ can be thought of as a 'spot' statistic. Very simple mathematical operations listed in Table 1 were chosen for the GP tree.

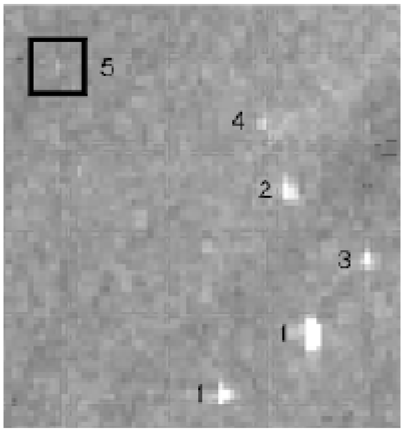

Figure 2: Presence of ships in the SAR image subjectively classified into grades.

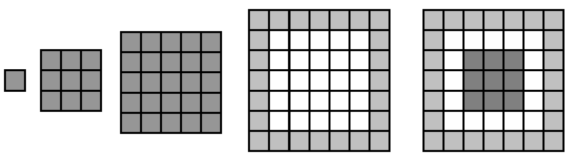

Figure 3: Examples of precomputed statistics defined in n x n (from left): pix, I_3, I_5, V_7 or P_7, and D_{3x7}.

3.1.1 Training data and fitness measures

The availability of the 50-m resolution SAR imagery made it practical to design a truth for the corresponding 100-m resolution SAR imagery. By this method it was possible to identify very faint objects in the 100-m resolution images. Therefore, objects in the truth varied quite significantly in size and in intensity making it possible to subjectively grade the ships in terms of their distinction. Five grades were arbitrarily determined, such that lower grades identified more distinct ships. For grade 5 sub-images, it was decided that the ships in the 50-m resolution sub-images were not distinguishable in the corresponding 100-m resolution sub-images. Referring to Figure 2 the grade 1 to 4 ships are clearly visible, but the grade 5 ship (in the box) is indistinguishable from the general ocean clutter. Consequently, only the clearer ships (grades 1 to 4) were used to evolve detectors.

The detector must both minimise the number of false alarms and the number of object misses. Fitness measures involving FP and TP were tested to determine which could most reliably produce accurate detectors. In other words, it was not an objective of this problem to set the balance between false alarms and misses to some precise level, but instead to consistently produce a family of detectors that could be considered superior in the Pareto ranking to those produced with other fitness measures or with rival techniques such as neural networks. While Pareto ranking is somewhat qualitative a comparison, the accuracy and generality of the evolved detectors was more exactly established by calculating a figure of merit (FOM) that assumed a neutral stance on desirability of false alarms and misses:

Table 1: GP parameter table

parameter	setting
functions	+, -, *, %, min, max.
terminals	integer (-127...128).
	real (0.0, 0.005... 1.0).
	real (0.0,-0.005...-1.0).
	pixel statistics: pix,
	$I_3...I_9, P_3...P_9, V_3...V_9$,
	$D_{3x5}, D_{3x7}, D_{3x9}$.
1st GP stage	5000 non-objects + 59 objects.
2nd GP stage	FP from 1st stage + 59 objects.

$$FOM = \frac{TP}{ships + FP} \tag{1}$$

whereby $ships$ stands for the total number of ship objects in the truth (including grade 5 ships).

The fitness measure that consistently produced superior detectors took ship grade into account by rewarding more highly for detections of the more distinguishable, lower grade, ship objects, and had the following form:

$$f_M = \frac{\sum_{hits}(5 - SG)}{\sum_{ships}(5 - SG) + FP} - 1 \tag{2}$$

Here $ships$ is the number of grade 1-4 ships in the training image, $hits$ is TP or the total number of detections and SG is the ship grade of the object in the truth.

Each evolutionary stage processed 59 ships (grade 1 to 4) from the training image. Both first and second GP evolution stages used the same fitness measure. The number of random non-object pixels (oceans) used in the first stage was 5000. Parameters such as population size, tournament size, and mating radius were varied and each combination was run using 20 random seeds (20 independent runs). Parameters were varied as follows: population sizes from 200 to 5000, tournament sizes from 2 to 16 and mating radiuses from 200 to population size. The random seed governed the various random selections, e.g. initial population, tournament selection, crossover point selection, etc. However, each run used the same 5000 non-object pixels.

As discussed in section 2, the maximum size of the detector in the first evolution stage was restricted to produce simple and fast detectors. Initially, the maximum tree size was varied. The same solution was derived whenever the size was greater than 100 nodes. The evolved detector always had less than 50 nodes and it commonly had less than 20 nodes. In order to encourage a simple solution, the maximum size of the first detector was set to 20 nodes. In addition, the first detector processed a minimal number of pixel statistics. Three statistics were first investigated: I_3, P_7 and V_7. However, it was found that *pix* or the value of the centre pixel was also required in order to hit small, faint ships. The first detector was allowed to evolve for exactly 10 generations. The maximum size of the detector of the second GP stage was set to 1000. Crossover was used in an elitist strategy with no mutation. The child was formed using the shorter side of the swap when crossover threatened to exceed the maximum prescribed tree size. The population of second detectors was evolved for 30 generations.

A procedure was developed to assess generalisation. For each run corresponding to a choice of GP parameters the ten best unique detectors at each generation of the second evolution stage were saved and then each of these was fused with the first detector. The FOM was computed over both test images, and the chosen fused detector was the one with the highest averaged FOM, provided that $FOM > 0.5$ for both images. As expected FOM always increased for the training image, but over-training occured after a given generation when the FOM begins to drop for some or all of images in the test set.

3.2 Discussion of Results

Conclusions of the GP parameter optimisation study can be found in [2-4]. This showed that the larger population and tournaments sizes more consistently produced superior ship detectors.

The performance of the best fused detector pair was measured up to results by Foulkes at DERA ([6] pg. 21) who tested a number of algorithms including multi-layer perceptron neural networks. His best results were obtained with a Kohonen SOM. These are displayed side by side to those with the best GP detectors in Table 2. The comparison is fair in the sense that both teams devoted a considerable time to this problem and used the same image for training the automatic detectors. From Table 2 there is no case where the GP method has fewer TP and more FP. It can be concluded that the two-stage evolution GP strategy evolved accurate ship detectors. It was also notable

Table 2: two-stage GP vs. SOM for all the images.

ship truth	SOM		GP	
grades 1-5	TP	FP	TP	FP
77	44	2	56	0
33	18	1	22	1
55	20	8	22	1
71	55	11	48	1
58	46	6	41	1

that certain random seeds resulted in small second detectors. For example, the performance of the following very short fit detector pair approached that of the champion detector:

1st detector:

$$\mathtt{min[pix - 0.49, max(pix, P_7)] - 0.325 - 1.05P_7 - V_7}$$

2nd detector:

$$\mathtt{pix - 1 - I_7 - V_7/0.285}$$

The second detector can be regarded as a method for reducing FP or number of false alarms at the expense of a slight increase in FN or number of missed objects. And when viewed this way it became interesting to investigate whether the effect was to simply move the threshold of detection from the condition: is object when $E > 0$ to the condition: is object when $E > \epsilon$, where ϵ is a small positive number and E stands for the evaluation of the detector function over the support n x n. But this was not to be the case - it was a more complex effect to achieve better detection rates [3].

3.3 Analysis of Detectors

Formulas for all champion detectors are reported in [3]. When simplified, the first detectors to a first approximation, had the following type of formulation:

pixel value - local mean - local variance - constant

This is equivalent to local spot detection and, in the absence of further information sources - such as intuition, past experience, geographical and scenario knowledge by analysts - could be considered to be a rough approximation to the ‘human eye’ approach to ship detection. Detectors evolved for the 50-m resolution images mixed a threshold type detector with a spot type detector - this is understandable when considering that ships in the 50-m resolution SAR images are clearer.

4 Multi-stage GP

This section demonstrates the utility of extending the two-stage GP to a strategy with multiple stages for the purpose of addressing the more challenging problem of object recognition in poorly constrained environments and with objects having large variability. It gives a very brief account of those experiments while [7,8] provide a more complete presentation.

4.1 Test problem

The objective was recognition of any motorised vehicle in infrared linescan (IRLS) imagery obtained by low flying missions. The objects varied in size in a variable environment, i.e. automobiles or lorries, as distinguished from roads, vegetation, and buildings. Figure 4 illustrates typical infrared signature for such vehicles and their large variability. The problem was made more difficult by its operational requirements. The detector must work with images taken at different heights and the performance of sensor equipment does in fact vary substantially.

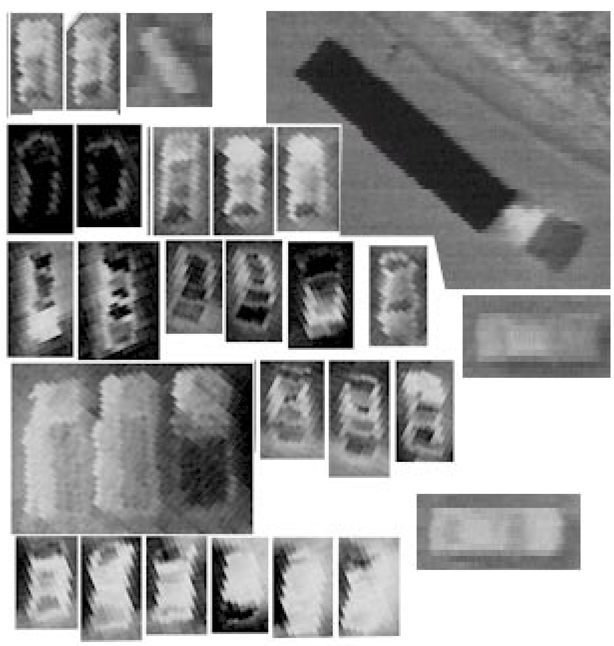

Figure 4: Examples of vehicles from the IRLS imagery

4.2 Formulation

As the vehicles vary in shape, size, orientation and appearance it was undesirable to detect vehicle features such as windscreens, bumpers, or wheels. The staged evolution strategy was designed to 'discover' these image features for itself with the first evolution stage tuned for sensitivity and the second evolution stage tuned for specificity. The n x n for these vehicle objects is $O(10^2)$ so that they cannot be represented by a single pixel.

A successful general scheme was developed. A vehicle was simply represented by a single pixel subjectively placed towards the center of the vehicle. A 'vehicle box' was centered on this point to contain most of the vehicle. GP was rewarded whenever it produced a detection within this box. The target had to be 'hit'.

A line was drawn along the major axis of each vehicle in the truth. Pixels on these lines were taken as the object pixels for the first evolution stage. Next, when sweeping the training images with the champion first detector, positives inside vehicle boxes were counted as object pixels for the second GP stage(s).

The second GP stage was decomposed into a number of sub-stages, e.g. 2i, 2ii, 2iii. The first second stage, 2i, attempted to eliminate all of the FP and to hit as many vehicles as possible. A second second stage, 2ii, took all misses from the 2i stage for its target pixels and tried to hit as many of them while again eliminating all of the FP. Subsequent second stages continued with this process until all targets had been hit.

Statistics were very similar to those in section 3, with the exception that the statistics of Figure 3 were made irrotational by defining them over concentric circles of thickness one pixel. Four types of statistics were defined on each circle at perimeters 11, 19, 27, and 35: (a) perimeter averages; perimeter standard deviations; edges found on the perimeter; and edge distribution norm values (reference [7,8] for definitions).

This choice of statistics provided the evolutionary process with: (1) an irrotational character; (2) robustness to high variability in pixel data; (3) both angular and radial information. Fourier analysis and invariant moments have only marginally improved the accuracy of the detectors - noisy imagery. The size of concentric circles was scaled according to flight height. The fitness measure was a version of equation (2) with constant $SG = 4$.

4.3 Results

Upon completion of the first stage, all detection within vehicle boxes were defined as vehicle points for the second evolution stage. This resulted in a larger coverage over vehicle areas, and many more vehicle boxes. GP then sacrificed many of these as the function of the second evolution stage is to trade off vehicle detection against false alarms. Vehicle pixels that survived this process were to be found in parts of the motorized vehicle that GP considered to best represent the vehicle. Therefore GP was able to discover both the characteristic features and their location by generalizing across these vehicle sub-images. This procedure overcame the need to a-priori discover features from vehicle sub-images that characterize all vehicles. Figure 5 illustrates this. It applies the evolved detectors to one of the images in the training set. Note that a second detectors may hit a car though it may have not trained on it.

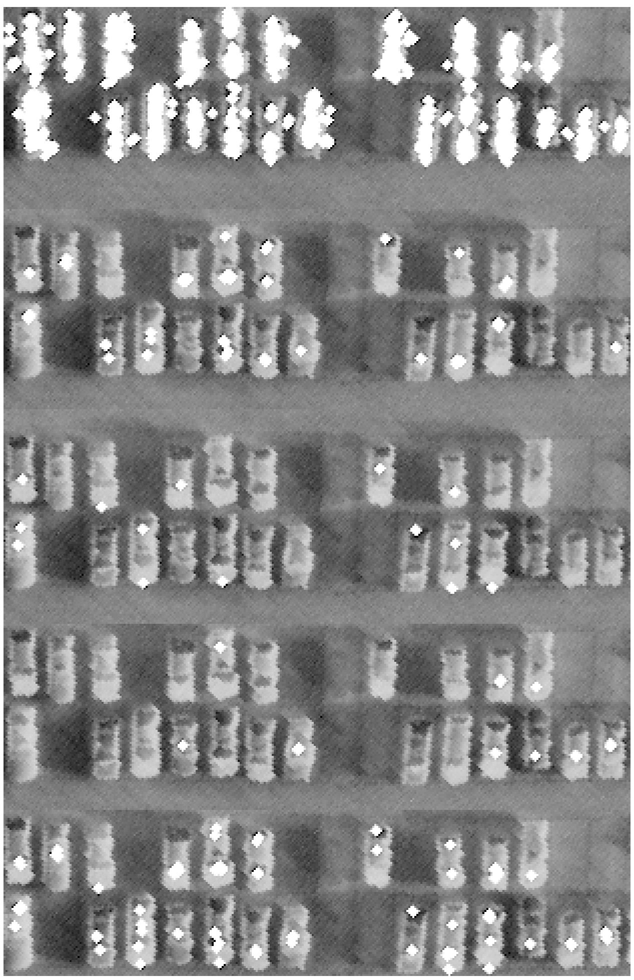

Figure 5: Illustration of multi-staged evolution strategy. From top to bottom: 1st, 1st+2i, 1st+2ii, 1st+2iii, fusion or 1st *AND* (2i *OR* 2ii *OR* 2iii)

The traditional method to handle infrared imagery has been to apply a simple filter that returns anything that is bright. The method we have developed is more powerful as it returns anything that looks like a vehicle. However, it is crucial to the stability-plasticity dilemma that the first evolution stage return a positive detection for any new vehicle that is presented for training. So far it always has. New and different types of vehicles are handled with a new second detector stage and the resulting detector is fused to the existing set with the *OR* function. The penalty in on-line evaluation of this approach is not significant.

5 Conclusions

The two-stage method has been shown to produce efficient and accurate detectors that generalise and compare well to those produced with rival AI techniques. The two-staged method is able to circumvent the need to control bloating in Genetic Programming.

A multi-stage method has been developed that does not require vehicle features to be prescribed a-priori. It is also able to address the so called stability-plasticity dilemma of inductive learners.

References

[1] Grossberg, S., *Studies of Mind and Brain*, Reidel, Boston, 1982.

[2] Howard D., Roberts S.C., and Brankin R., Advances in Engineering Software, 30 (303-311), 1999.

[3] Roberts S.C., Howard D., and Brankin R., *Genetic evolution of automatic ship detection in SAR imagery.* DERA Malvern technical report DERA/CIS/SEC/TR980323, 1998.

[4] Howard, D., Roberts S.C., and Brankin R.,*Evolution of Ship Detectors for Satellite SAR Imagery*, EuroGP'99, pp. 135-148, Göteborg, Sweden, May 1999, Springer LNCS.

[5] Poli, R., *Genetic Programming for Image Analysis.* In Koza, John R., Goldberg, David E., Fogel, David B., and Riolo, Rick L. (editors). Genetic Programming 1996: Proceedings of the First Annual Conference Cambridge, MA: The MIT Press.

[6] Foulkes, S.B. *Ship detection in ERS-1 and Radarsat SAR images using a self-organising neural network.* DERA Malvern technical report DERA/LS1/TR980309, 1998.

[7] Roberts S.C., and Howard D., *Evolution of Vehicle Detectors for Infrared Line Scan Imagery*, EvoIASP'99 and EuroEcTel'99 Joint Proceedings, pp. 111-125, Göteborg, Sweden, May 1999, Springer LNCS.

[8] Howard D., and Roberts S.C., EUROGEN'99 Lecture Notes, Short Course on Evolutionary Algorithms in Engineering and Computer Science, University of Jyväskylä, Finland, May 1999.