

A Multi-Species Genetic Algorithm Applied to Radar Scattering Centre Identification in Three-Dimensions

Evan J. Hughes
hughesej@rmcs.cranfield.ac.uk

Maurice Leyland

Brian A. White

School of Engineering and Applied Science,
Cranfield University,
Royal Military College of Science,
Shrivenham, Swindon, SN6 8LA.

Abstract

Realistic radar target models are required for use in a missile–target simulation system. In this paper a Multi-Species Genetic Algorithm is used to extract a radar cross section model from high-resolution three-dimensional radar target images by identifying the dominant scattering centres. A typical image may require in excess of 2 Gbytes of storage space. Data retrieval overheads on small systems render iterative methods of model extraction impractical. This method exploits the niche-finding characteristics of the multiple species in the genetic algorithm to simultaneously identify multiple scattering centres. This significantly reduces the data access requirements with only a moderate increase in processing overhead.

I Introduction

The extracted models created by the algorithm detailed in this paper are designed to be used with a synthetic missile in 3-dimensional virtual engagement scenarios. Real high-resolution target radar cross section data consumes massive amounts of disk storage space. This requirement may be overcome by identifying and storing the locations of the main reflective surfaces of the target. These collections of scattering centres can be replaced by models consisting of ideal isotropic point scatterers placed in the same locations. This sufficiently approximates the target for our purposes. Models generated from many different target aspects may be combined with synthetic details, creating a realistic radar cross section with 4π steradian coverage. The locations of the reflective surfaces may be found by analysing Inverse Synthetic Aperture Radar (ISAR) images of the target. These images are the radar equivalent of an optical hologram. An iterative method is often used to identify

the scatterer locations [Bhalla and Ling, 1996]. This operates by first finding the size and coordinates of the brightest spot in the image and places a scatterer in the corresponding position in the model. An ISAR image of the model is produced and subtracted from the original image to remove the corresponding bright spot. The process is repeated until all bright spots of interest have been removed. Other methods have been developed that rely on drawing contour maps of the image and locating the scatterers within the bounded regions [Breuille and Caille, 1994].

These methods work well but have one major drawback: A typical three dimensional image of 512 pixels on each axis will require $(512)^3$ complex numbers and therefore two Giga-bytes of storage space. Finding the location of the maximum value necessitates searching the entire data set for each scatterer that is resolved. Images often require 100 or more scatterers for accurate representation and therefore the equivalent of 200 Gbytes of data must be retrieved from the storage media. On a small system, the data access and transfer times are significant when compared to the processing overhead. Producing a contour map of the image is more intensive and therefore impractical.

A solution is therefore proposed where a genetic algorithm is used to locate multiple bright spots in one pass. These can then be formed into a model and its effects subtracted from the original image as before. Further applications of the GA will locate any smaller points remaining. This multi-modal function approach, despite the small increase in processing overhead, can make model calculation viable on a small system.

II Multi-Modal Optimisation and Sharing

Conventional genetic algorithms use a single population of a single species. The algorithms are designed

so the solutions presented by the different individuals converge on the single optimum solution of the objective function. In multi-modal optimisation, the genetic algorithm is designed to converge with multiple solutions, each corresponding to a separate peak in the objective function. A sharing mechanism may be used to force a genetic algorithm to exhibit multi-modal behavior [Goldberg, 1989, Pages 185–197]. The sharing system operates by modifying the objective value that is seen by each individual. If a number of individuals all occupy the same peak in the objective function they are made to share the objective value at that point. This simple concept is enough to allow multiple stable populations to form.

In a practical algorithm, a sharing function that is related to the separation distance between two individuals (genotypic or phenotypic space) is used to control the modification of the objective function. Equation 2 [Michalewicz, 1992, Page 176] defines the sharing function used, with $d(\chi(i), \chi(j))$ defined as the distance between the chromosomes $\chi(i)$ and $\chi(j)$, $s(\chi(i), \chi(j))$ is the sharing effect of j on i and α is a factor for modifying the function shape. This function produces a linear result that moves from unity at zero distance to $1 - \alpha$ at a distance of σ and then zero thereafter.

$$d = d(\chi(i), \chi(j)) = |\chi(i) - \chi(j)| \quad (1)$$

$$s(\chi(i), \chi(j)) = \begin{cases} 1 - \left(\frac{d(\chi(i), \chi(j))}{\sigma}\right) \alpha & d \leq \sigma \\ 0 & d > \sigma \end{cases} \quad (2)$$

For each individual chromosome, $\chi(i)$, the distance is calculated to every other chromosome, $\chi(j)$, in a population of N individuals and the values for each of the sharing functions are totalled (equation 3). The result is used to derate the raw objective value $\text{obj}(\chi(i))$ yielding a new objective value $\mathcal{O}(\chi(i))$. This is shown in equation 4.

$$\mathcal{S}(\chi(i)) = \sum_{j=1}^N s(\chi(i), \chi(j)) \quad (3)$$

$$\mathcal{O}(\chi(i)) = \frac{\text{obj}(\chi(i))}{\mathcal{S}(\chi(i))} \quad (4)$$

These sharing functions work well but crossover between individuals from different populations often yields poor solutions and therefore convergence is slow and unpredictable. The individuals have a tendency to distribute themselves evenly through out the objective surface. This leads to a large proportion of the population being located on the highest peaks and smaller peaks holding proportionally less individuals. For the large and complex optimisation surfaces found in scattering centre identification, large populations

are required. Thus the requirement for every individual to be compared to every other produces a significant processing overhead.

To improve the effects of the crossover operations, a selective breeding strategy may be employed. This effectively only breeds like-with-like and therefore crossover tends to produce better offspring [Goldberg, 1989, Page 192–197]. This is easily accomplished by splitting the main population into a number of subpopulations that act as independent mating pools. If the sharing strategy is applied as before, the multi-species algorithm converges faster. The number of peaks that are found is often restricted to the number of species though but a smaller total population may be used to achieve similar results.

Due to the nature of the sharing function, absolute convergence of all similar individuals on one point is unlikely. This leads to the peak of the optima being poorly identified. If we add an additional sharing rule that says an individual only has to share with members of other species, each species will converge on the peak of an optima, with many individuals identifying the true peak value. This effect may be further enhanced by controlling the size of genetic mutations. By reducing the size of the mutations with each subsequent generation, each species will eventually be confined to its own locality and be forced to perform a fine search of the area.

III Modified Multi-Species Algorithm

The sharing functions work well but suffer from a severe processing overhead with every individual having to be compared with every other. In an attempt to reduce this, the sharing function has been modified to operate with the statistics of each species rather than its members. We can define the position and spread of a species by the mean of the species chromosomes and their standard deviation. Equations 5 and 6 define this, where n_s is the number of individuals in a species and $\chi(i, k)$ denotes the chromosome of individual i of species k . If we assume that the spread of individuals around the mean position is roughly Gaussian, a sphere with a one standard deviation radius from the mean will encompass 68% of the population. A two standard deviation sphere will cover 95%. Thus we may define $2\sigma_k$ as covering 95% of population k .

$$\bar{\chi}_k = \frac{1}{n_s} \sum_{i=1}^{n_s} \chi(i, k) \quad (5)$$

$$\sigma_k^2 = \frac{1}{n_s} \sum_{i=1}^{n_s} |\chi(i, k) - \bar{\chi}_k|^2$$

$$= \frac{1}{n_s} \sum_{i=1}^{n_s} d(\chi(i, k), \bar{\chi}_k)^2 \quad (6)$$

Equation 7 defines the modified sharing function, where σ' is a minimum sharing distance analogous to σ in equation 2. This sharing function is then applied to all N_s species, except the members own, and the results summed (equation 8). The objective cost for the individual is then derated by one plus the share value to account for the individual itself and is shown in equation 9.

$$s(\chi(i, j), \bar{\chi}_k) = \begin{cases} 1 - \left(\frac{d(\chi(i, j), \bar{\chi}_k)}{\sigma'_k} \right)^2 \alpha & d \leq \sigma'_k \\ 0 & d > \sigma'_k \end{cases} \quad (7)$$

$$\text{where } \sigma'_k = \begin{cases} \sigma_k & \sigma_k > \sigma' \\ \sigma' & \sigma_k \leq \sigma' \end{cases}$$

$$\mathcal{S}(\chi(i, j)) = n_s \sum_{k=1}^{N_s} \frac{s(\chi(i, j), \bar{\chi}_k)}{2\sigma_k} \Big|_{k \neq j} \quad (8)$$

$$\mathcal{O}(\chi(i, j)) = \frac{\text{obj}(\chi(i, j))}{1 + \mathcal{S}(\chi(i, j))} \quad (9)$$

The sharing function defined in equation 7 is based on a squared law rather than a linear function. This affects individuals that are close to the species centre more and allows an increase in calculation speed by not requiring the square root of the magnitude of the distance to be taken. Unlike the fixed shape individual sharing functions used previously, the functions associated with each species are dynamic and vary with the geographical motion of the individuals within the species. Figure 1 depicts the function shape graphically for a number of different species spreads. The share values plotted on the y-axis are the results for equations 7 and 8 for a single species with 10 individuals and the conditions for α (alpha), σ (sigma) and σ' (sigma') shown on the graph. When a species population is widely dispersed, the function has little effect on other individuals. As a species population converges, the range of the function decreases but its influence increases. This forces different species to separate as their populations converge. A minimum distance, σ' , for the spread of the function is used to prevent different species from converging too close to one another. This helps increase the diversity of the geographical spread of the species.

By calculating the distance to each species mean position rather than to each individual, a significant speed increase is achieved with the modified algorithm. Experiments appear to show that the overall performance of the algorithm is not compromised by the changes.

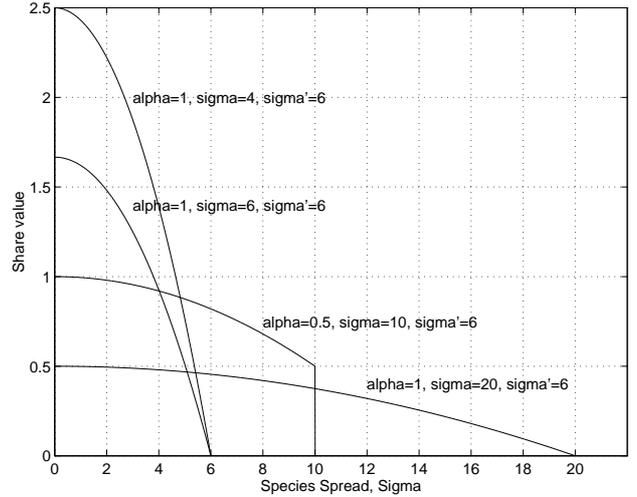


Figure 1: Typical Sharing Function Shapes

IV Algorithm Construction

A real valued chromosome with three parameters has been used to define each individual. The parameters are defined as being the (x, y, z) coordinates of a location in the three-dimensional ISAR image. The genotypic level allows the parameters to have fractional components. The genes are rounded to the nearest integer to obtain the phenotypic data. The raw objective value is defined as the image intensity at the indexed point.

The algorithm follows the usual format of ranking, selection, crossover, mutation and evaluation but with each species being processed separately. The same number of offspring are generated as parents and a total replacement policy is used. This helps slow convergence and allows the species to relocate themselves to minimise problems caused by overcrowding.

The fitness value $F(x)$ is assigned according to rank position p_x of individual x . The individual with the lowest $\mathcal{O}(x)$ (least fit) being assigned a rank position of 1 and the best individual being assigned rank position M . Equation 10 details the calculation of $F(x)$.

$$F(x) = \frac{2(p_x - 1)}{M - 1} \quad (10)$$

Stochastic Universal Sampling [Michalewicz, 1992, Page 57] is used to select M individuals from the population, each individual having a probability of selection defined in equation 11.

$$\text{Prob}(x \text{ selected}) = \frac{F(x)}{\sum_{i=1}^M F(i)} \quad (11)$$

The individuals selected are randomly shuffled and then paired up for breeding. Uniform Crossover [Michalewicz, 1992, Page 88] is used to generate two new offspring from each pair of parents. This operator simply swaps over the genes of the parents with a probability of 0.5. This crude operation on a real-valued chromosome reduces the convergence speed and helps to maintain a good geographical spread of the individuals.

Genes are mutated with a probability of 0.3 with the range of the mutation being governed by equation 12. This *Non-Uniform* mutation [Michalewicz, 1992, Chapter 6] function is unity initially and progresses to zero at the final generation and is used to modify the maximum deviation from the current gene value, where G is the generation number and G_m is the maximum number of generations. Initially, the gene can mutate to any value within its constrained range but this is reduced with time. The range modifier function forces the genetic algorithm to converge on a solution by confining the offspring of each subsequent generation to a diminishing region.

$$R(G) = 1 - \left(\frac{1 - \cos\left(\frac{(G-1)\pi}{(G_m-1)}\right)}{2} \right)^2 \quad (12)$$

Figure 2 depicts the function shape graphically.

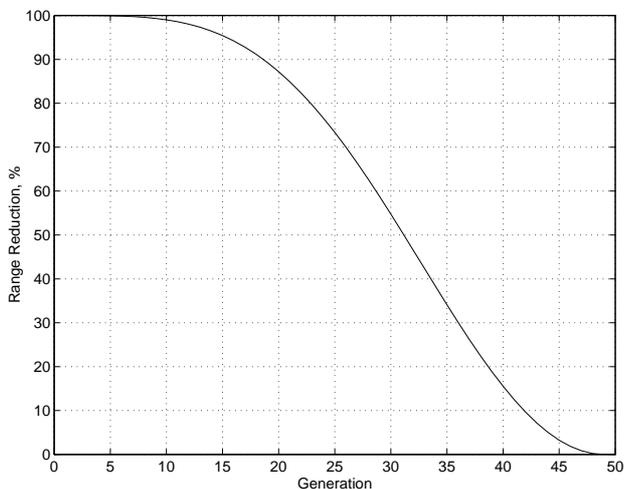


Figure 2: Mutation Range Modifier Function

Various function shapes have been tried including linear, square law and raised cosine. The nature of the species behavior is that initially, the uniform distribution of individuals leads to the species mean positions all being located at the geographic centre of

the image. During the first few generations, the natural instinct of all the species is to colonise the highest peak. By around the fifth generation, the main peak becomes crowded and the species are forced to diversify and find other smaller niches to colonise. This diversification requirement means that the available mutation range must be large for the first half of the generations to allow the species to explore the image properly. This natural crowding tendency is also reduced by impeding the natural convergence of the GA through the use of the crude crossover mechanism and total replacement. In the final stages of the GA, the mutation range is limited to a very small locality, forcing the species to converge on the true local optimum.

The objective function for the offspring is calculated as previously detailed based on the statistics of the parent population. Policies other than total replacement have been tried but appear to offer little benefit as the objective function is effectively dynamic with the motion of the species.

The GA is terminated after 50 generations and the best overall individual of each species is recorded as a solution. Trials have shown that many species each with small populations produce better results than a few species consisting of large numbers of individuals. This is due to the small species being able to adapt more quickly to the dynamic objective function. A small number of large species will all try to colonise the highest peak only.

Experiments have shown that 30 species each with a population of 15 individuals provide good results over a range of different images. Over 50 generations, this gives 22500 objective calculations for each run of the genetic algorithm and can locate up to 30 peaks.

V Results

For the purposes of presentation, the following results are for two-dimensional images. All the previous theories and algorithms apply except that the z coordinate is forced to always be zero.

The performance and current state of the algorithm may be assessed using the mean of the standard deviations of the species chromosomes. This gives an indication of the geographical spread of the species individuals. The upper section of figure 3 shows the trace of the mean spread of 30 species. With the multi-species genetic algorithm, the objective functions of the species vary dynamically with the species motion. The lower portion of figure 3 shows the best species mean objective for each generation.

The objective functions yield little reliable information to describe the current state of the algorithm. Thus algorithm convergence is determined by the con-

vergence of the species chromosomes rather than the convergence of the objective function as in a conventional genetic algorithm. To test the algorithm, a simple model with nine scattering centres was used to generate a low-resolution test image. A genetic algorithm consisting of 30 species, each with a population of 15 individuals was applied to the image. The algorithm ran for 50 generations with a sharing limit of $\sigma' = 3$ and $\alpha = 0.5$ (equation 7).

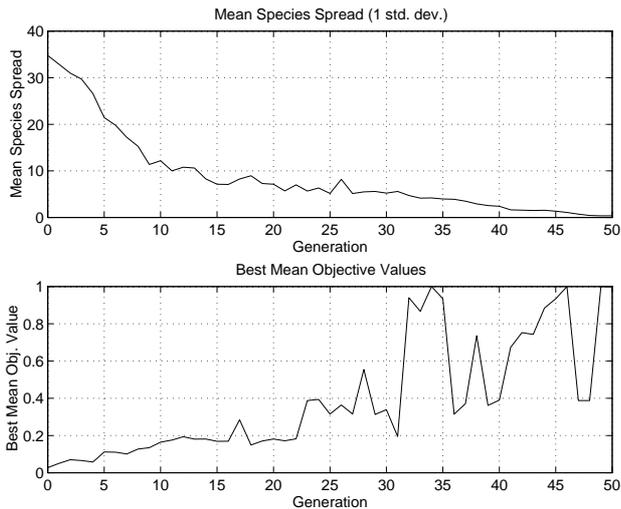


Figure 3: Algorithm Statistics

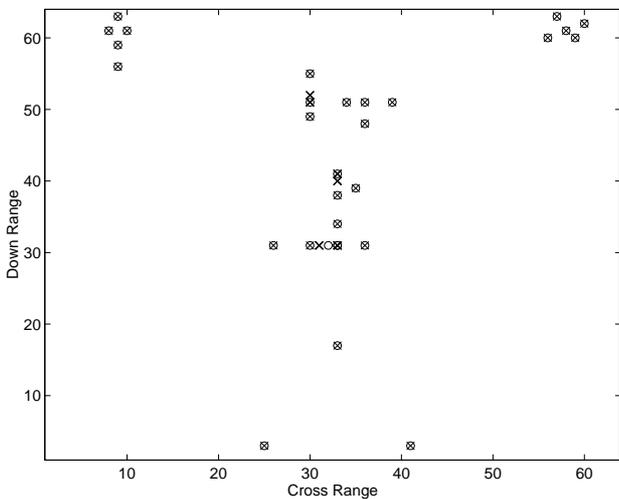


Figure 4: Final Geographical Spread of Species

Figure 4 shows the geographical locations of the species after the final generation. Each individual is represented by a cross and each species centre is represented by a circle. It is easily seen that for the majority of the species, convergence is total. Only 29

species are visible on the graph as the main peak in this example is supporting two species. This is a rare occurrence. Figure 5 shows the nine identified peak locations plotted on the objective surface. These points were extracted from the species results by identifying all the species whose best objective value had a relative amplitude greater than 0.1. Any duplicated points are removed and the resulting positions plotted. It is quite obvious from figure 5 that the genetic algorithm had no difficulty in identifying the peaks.

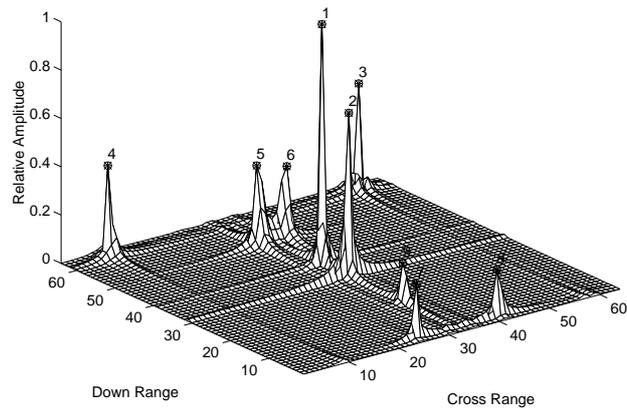


Figure 5: Surface Plot of Image with Identified Peaks

Figures 6 and 7 show the results of the above algorithm applied to a high-resolution ISAR image. The image in figure 6 is of a 110 point test model with a wide range of scatterer amplitudes. After five applications of a 30-species GA, 104 points have been identified. The equivalent ISAR image of the model is shown for comparison in figure 7.

Applying the algorithm to three-dimensional images poses no problems, with good results being obtained with between 30 and 50 species which have from 10 to 15 individuals per species. trials suggest that with more than 50 species in an algorithm, overcrowding can occur with multiple species occupying the same peak. This can be counterproductive and it is better to use fewer species and re-apply the GA instead.

VI Conclusions

The modified multi-species genetic algorithm described in this paper has proved to have excellent niche-finding properties for solving multi-modal optimisation problems. The reduced data retrieval requirements for large problems leads to a significant run-time saving over traditional algorithms where the

entire data set has to be searched multiple times. Experiment has shown that the GA approach can simultaneously identify 10 locations reliably with only 25000 data accesses. If we apply the GA 10 times to identify 100 scatterers, a total of 250,000 accesses are required. The iterative approach accesses over 500 times as much data to achieve the same model resolution. Two and three dimensional implementations have been tested. Extension to further dimensions for other problems should pose no difficulties.

With the current algorithm implementation, each species has an equal probability of being able to colonise the highest peak. This causes the species to crowd the highest peak initially and then diverge when one species has gained a foothold. From these observations of species behavior, it is hypothesised that by using species with different sized populations, a bias is introduced that will eventually lead to the largest species colonising the highest peak and the other species spread according to peak and population size. The smaller populations would be more dynamic and therefore more able to find uncolonised peaks. This should help maintain the genetic diversity in the algorithm and therefore improve the geographical spread of the obtained solutions. A typical spread of population sizes may be one large species with fifty individuals down to many species each with say four individuals.

Finally it should be noted that in this particular algorithm, limited crossover between species is not applied. A small amount of chromosome migration between species was tried but it was found that the performance of the algorithm was seriously reduced. The effect of the migration was to make all the species attempt to converge on the highest peak, causing remote peaks to be missed by the algorithm.

References

- [Bhalla and Ling, 1996] Bhalla, R. and Ling, H. (1996). Three-dimensional scattering centre extraction using the shooting and bouncing ray technique. *IEEE Transactions on Antennas and Propagation*, 44(11):1445–1453.
- [Breuille and Caille, 1994] Breuille, C. L. and Caille, E. (1994). Modeles multi-points brillants de signatures radar de cibles. In *Radar International Conference*, pages 325–329, Paris, France. Soc. Electr. & Electron.
- [Goldberg, 1989] Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley Publishing Company, Inc.
- [Michalewicz, 1992] Michalewicz, Z. (1992). *Genetic Algorithms + Data Structures = Evolution Programs*. Springer-Verlag, 2nd edition.

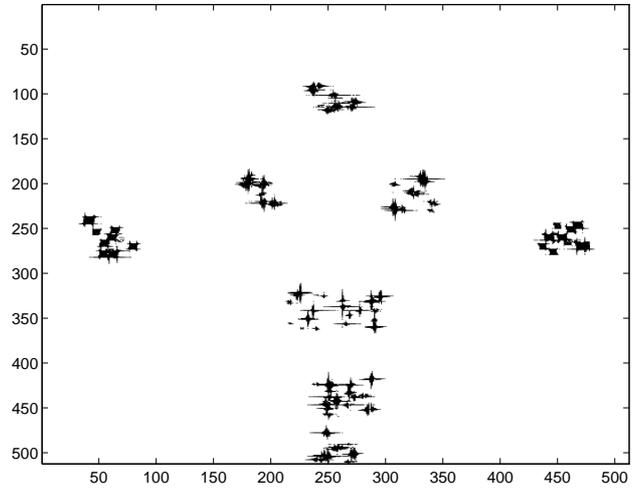


Figure 6: High-Resolution ISAR Plot, Target Data

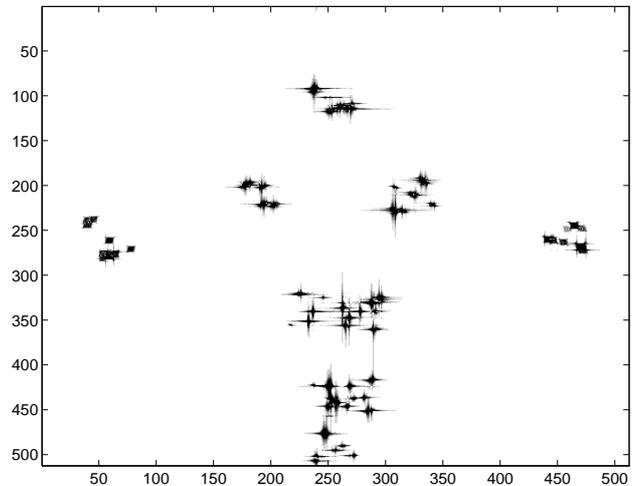


Figure 7: High-Resolution ISAR Plot, Model Data