Object Recognition by Flexible Template Matching using Genetic Algorithms [*]

A. Hill, C. J. Taylor and T. Cootes.

Department of Medical Biophysics, University of Manchester,
Oxford Road, Manchester M13 9PT, England.

Abstract. We demonstrate the use of a Genetic Algorithm (GA) to match a flexible template model to image evidence. The advantage of the GA is that plausible interpretations can be found in a relatively small number of trials; it is also possible to generate multiple distinct interpretation hypotheses. The method has been applied to the interpretation of ultrasound images of the heart and its performance has been assessed in quantitative terms.

1 Introduction

Flexible template matching has become an established technique for image interpretation [4,5,6]. It is particularly relevant to problems such as medical image interpretation, where there can be considerable variation between examples of the same object. The basis of the approach is to project instances of the template into the image until one which is well supported by the observed data is found. There are several requirements:

- A template with a (small) number of controlling parameters; for any particular set of parameter values it must be possible to reconstruct a feasible instance of the object the model represents.
- For any model instance which is projected back into the image an *objective function* which can assess support provided by the image evidence; in general this objective function is likely to be non–linear with respect to the model parameters, multi–modal, possibly noisy and/or discontinuous.
- A method of optimisation which can search the parameter space of the model in order to identify the particular set of parameter values for which the objective function is maximised i.e. for which the image evidence is greatest.

In this paper we have employed the approach described by Cootes et al [1] for constructing flexible template models which satisfy the first requirement listed above. For the echocardiogram exemplar we consider in this paper, the variation in shape can be described approximately by 6 parameters. In general, it is also necessary to translate scale and rotate the template onto the image, giving 4 additional parameters for 2D models. Altogether, then, we have 10 parameters which instantiate and project the model onto the image. If we employ n bits to encode each parameter a search space of $2^{10\,n} \approx 10^{3\,n}$ results. Clearly, even for modest values of n, the search space is large. In the literature this problem has been overcome by assuming that a reasonably good approximation to the solution is available so that the search space can be drastically reduced. We argue that this is often an unrealistic assumption and have attempted to develop methods of finding good solutions without restricting the search space.

The problem, as stated, is one of optimisation of a multi–modal, non–linear function of many variables, the function being possibly discontinuous and/or noisy. A class of methods which have been proposed as a solution to problems of this nature are Genetic Algorithms (GAs) [2,3]. It is claimed that GAs can robustly find good solutions (although not guaranteed to be optimal) in large search spaces using very few trials. Given an objective function, f, which measures the evidential support for any particular projection into the image of the model, a GA search can find a set of parameters which

---

provide a good explanation (or interpretation) of the image. Our results demonstrate the feasibility of this approach.

2 Genetic Algorithms

GAs employ mechanisms analogous to those involved in natural selection to conduct a search through a given parameter space for the global optimum of some objective function. The main features of the approach are as follows :

- A point in the search space is encoded as a *chromosome*.
- A *population* of N chromosomes/search points is maintained.
- New points are generated by probabilistically combining existing solutions.
- Optimal solutions are *evolved* by iteratively producing new *generations* of chromosomes using a *selective breeding* strategy based on the relative values of the objective function for the different members of the population.

A solution $\underline{x} = (x_1, x_2, .., x_n)$, where the x_i are in our case the model parameters, is encoded as a string of *genes* to form a *chromosome* representing an *individual*. In many applications the gene values are $[0,1]$ and the chromosomes are simply bit strings. An objective function, f, is supplied which can decode the chromosome and assign a *fitness value* to the individual the chromosome represents.

Given a population of chromosomes the genetic operators *crossover* and *mutation* can be applied in order to propagate *variation* within the population. Crossover takes two *parent* chromosomes, cuts them at some random gene/bit position and recombines the opposing sections to create two *children* e.g. crossing the chromosomes 010–11010 and 100–00101 at position 3–4 gives 010–00101 and 100–11010. Mutation is a background operator which selects a gene at random on a given individual and mutates the value for that gene (for bit strings the bit is complemented).

The search for an optimal solution starts with a randomly generated population of chromosomes; an iterative procedure is used to conduct the search. For each iteration a process of *selection* from the current *generation* of chromosomes is followed by application of the genetic operators and re-evaiuation of the resulting chromosomes. *Selection* allocates a number of trials to each individual according to its *relative fitness value* $f_i/\bar{f}$, $\bar{f} = 1/N \{f_1 + f_2 + .. + f_N\}$. The *fitter* an individual the more trials it will be allocated and vice versa. Average individuals are allocated only one trial.

Trials are conducted by applying the genetic operators (in particular crossover) to selected individuals, thus producing a new generation of chromosomes. The algorithm progresses by allocating, at each iteration, ever more trials to the high performance areas of the search space under the assumption that these areas are associated with short sub–sections of chromosomes which can be recombined using the random cut–and–mix of crossover to generate even better solutions.

3 Flexible Template Construction

We have employed a modified version of the method described by described by Cootes et al [1] for constructing flexible templates. The templates are generated from a set of examples of the object to be modelled. The technique is to represent each example by a set of labelled points (x_i, y_i) so that each (x_i, y_i) is at an equivalent position for each example. The mean position of the points gives the average shape of the template and a principle component analysis of the deviations from the mean gives a set of modes of variation. These modes represent the main ways in which the shape deforms from the mean. By adding weighted sums of the first m modes to the average shape, new examples of the object can be generated. The m weights $\underline{b} = (b_1, b_2, .., b_m)$ give a set of parameters for the model.

4 The Echocardiogram Exemplar

We have evaluated the method using ultrasound images of the heart (apical 4–chamber echocardiograms). A typical example is shown in figure 1.a. The problem we address
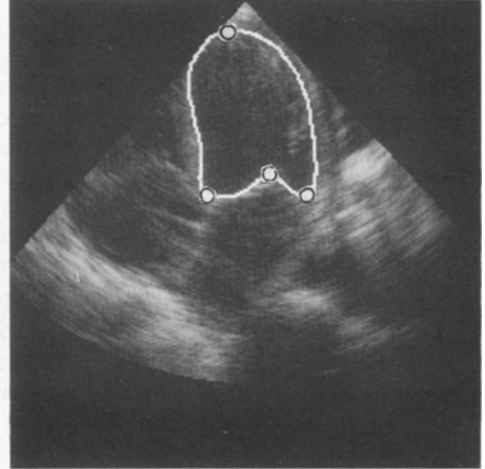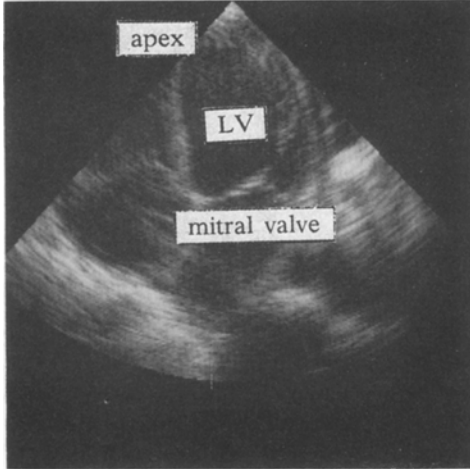

Figure 1.a : Example echocardiogram   *Figure 1.b : Associated LV boundary*

is that of locating the boundary of the left ventricle (LV) as shown in figure 1.b. The aim is to provide quantitative information concerning LV function.

In order to construct a flexible template model of the left ventricle we employed a training set of 66 apical 4–chamber echocardiograms with LV boundaries labelled by an expert. On each of these boundaries 4 fiducial points were placed by the expert at the apex and mitral valve – see figure 1.b. A number of intermediate points were generated automatically by positioning equidistant points along the boundary to give 18 points in all. The principle component analysis identified 6 major modes of variation in LV shape for this training set.

5 The Objective Function

To perform the experiments described below we used a simple and rather ad–hoc objective function. A number of points, P, on the boundary of the candidate object are selected for processing. For each of these points a grey level profile perpendicular to the boundary at that point is extracted. For each profile the position $(p_i, \ p_{min} \le p_i \le p_{max})$ and strength (g_i) of the largest intensity step along the profile are recorded. The objective function is then given by :

$$f = \frac{1}{\bar{g}} \frac{1}{P} \sum_{i=1}^{P} \left\{ \left| \frac{p_i}{p_{min}} \right| + \left| \frac{g_i}{\bar{g}} - 1 \right| \right\} \quad \text{where} \quad \bar{g} = \frac{1}{P} \sum_{i=1}^{P} g_i$$

When we use the objective function to evaluate instances of the model we seek to minimise f, thus favouring solutions with strong edges ($\bar{g}$ large) of equal magnitude $(|g_i/\bar{g} - 1| \to 0)$ located close to the boundary position predicted by the model $(p_i \to 0)$.

6 Results

In order to assess the method we employed 2 frames from each of ten echocardiogram time sequences. The frames showed the LV in its most extended and contracted state for each sequence and the LV boundary was delineated on each frame by an expert. The images were interpreted using the model-based approach described above and the

resulting LV boundaries were compared with the expert–generated boundaries using the mean pixel distance between boundaries $\Delta = \dfrac{1}{P} \sum\limits_{i=1}^{P} \sqrt{(a_{i,x} - b_{j,x})^2 + (a_{i,y} - b_{j,y})^2}$ where $\underline{b}_j = (b_{j,x}, b_{j,y})$ is the point on the expert boundary closest to the point $\underline{a}_i$ on the GA generated boundary. The control parameters for the GA were : population size (N) = 50, crossover rate (C) = 0.6 and mutation rate (M) of 0.005. A limit of 5000 objective function evaluations was employed. For 19 of the 20 images $\Delta \le 6$ and for 14 of the 20 images $\Delta \le 4$. A typical result is show in figure 2.
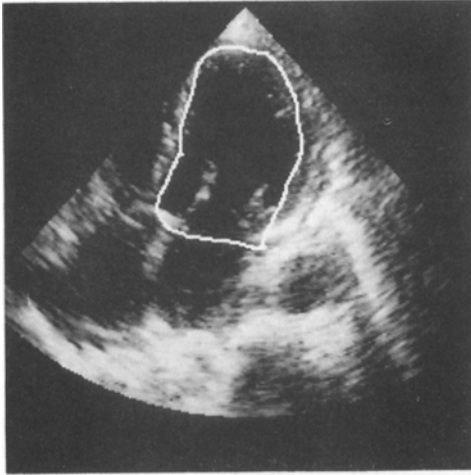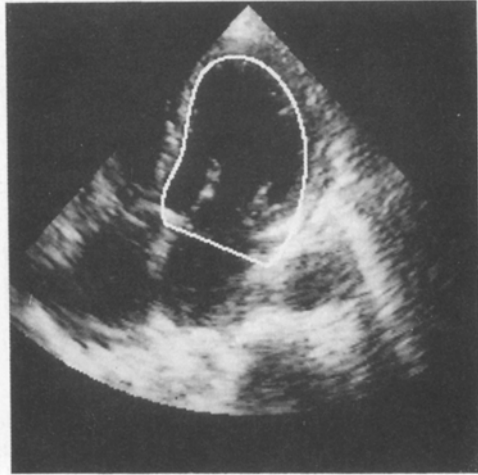

Figure 2.a : Expert boundary.   Figure 2.b : GA boundary.

7 Niches and Species

During the course of the above experiments it became obvious that one of the major problems for the GA search was premature convergence to sub–optimal solutions in the presence of multiple plausible interpretations. Rather than extract a single solution, we would like the GA search to extract a handful of strong, ranked candidates for the object we wish to locate. The problem of locating multiple optima when using GAs is discussed by Goldberg [2]. The approach adopted is to reduce the number of individuals in over–crowded areas of the search space by modifying their function values.

In this case the fitness of an individual is *weighted* by the number of *neighbours* the individual has. The more neighbours an individual has the worse its fitness value. The number of individuals allowed to crowd into the various areas of the search space is thus proportional to the relative fitness value associated with the different areas of the search space. In order to maintain stable sub–populations two further modifications are necessary. The first is simply to increase the size of the population in order to prevent extinction due to sampling errors. The second is to implement a restricted mating strategy in order to promote speciation i.e. prefer neighbours to distant individuals for crossover. The key to the entire procedure is the ability to decide *Who is a neighbour?* i.e. how *close* are two points in the search space. We define two individuals $\underline{x}_1, \underline{x}_2$ to be neighbours if they lie within a sub–space defined by the global transformation parameters, translation (t_x, t_y), scaling (s) and rotation (ϕ) :

$$|t_{x,1} - t_{x,2}| \le \delta t_x , \quad |t_{y,1} - t_{y,2}| \le \delta t_y , \quad |s_1 - s_2| \le \delta s , \quad |\phi_1 - \phi_2| \le \delta\phi .$$

We have employed this mechanism to extract multiple solutions from echocardiogram images. An example of applying the method is shown in figure 3. The groups are listed in order of size, group 1 containing the largest number of individuals. The *species* these groups represent were generated using $N = 100$, $C = 0.5$, $M = 0.005$ and 100 generations.
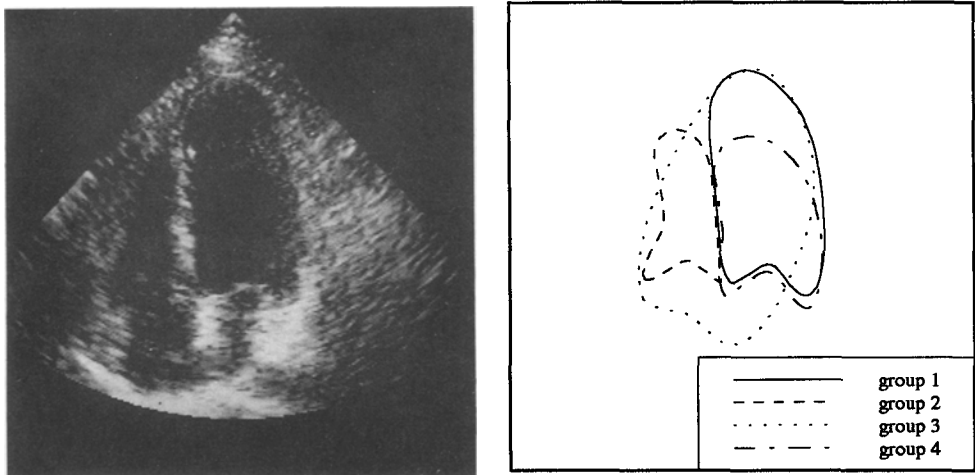

Figure 3 : Species adapted to different chambers.

8 Conclusions

One of the major attributes of the model–based approach is the ability to correctly interpret incomplete and/or noisy image data by constraining all possible interpretations using knowledge represented by a model. In order for this process to be successful it may be necessary to search a high–dimensional, non–linear, multi–modal and noisy search space. The combination of a flexible template model and a powerful search technique has been shown to give very promising results. The major feature of GAs which is useful for object location is a population of solutions. This allows alternative interpretations to compete with one another, the strongest solution having the greatest probability of success. This property of a GA search also allows multiple plausible interpretations to be extracted.

References

[1] Cootes T. F., Cooper D. H., Taylor C. J., Graham J. : A Trainable Method of Parametric Shape Description. Proc. British Machine Vision Conference, Glasgow (1991) Springer Verlag 54–61.

[2] Goldberg D. E. : Genetic Algorithms in Search, Optimisation and Machine Learning. Addison–Wesley (1989).

[3] Holland J. H. : Adaptation in Natural and Artificial Systems. University of Michegan Press, Ann Arbor (1975).

[4] Lipson P., Yuille A.L., O'Keeffe D., Cavanaugh J., Taaffe J., Rosenthal D. : Deformable Templates for Feature Extraction from Medical Images. Proc. 1st European Conference on Computer Vision, Lecture Notes in Computer Science, Springer–Verlag, (1990) 413–417.

[5] Pentland A., Sclaroff S. : Closed–Form Solutions for Physically Based Modeling and Recognition. IEEE Trans. on Pattern Analysis and Machine Intelligence 13(7) (1991) 715–729.

[6] Yuille A.L., Cohen D.S., Hallinan P. : Feature Extraction from Faces using Deformable Templates. Proc. Computer Vision, San Diego (1989) 104–109.