

A Distributed Approach to Image Interpretation Using Model-Based Spatial Reasoning

A.Ratter, O.Baujard, C.J.Taylor and T.F. Cootes

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

ABSTRACT

We address the problem of finding a consistent interpretation of an image when a number of object features may be detected independently, but unreliably, and their relative positions are known to be constrained. Our method treats feature detection and the application of spatial constraints as co-operating processes. We show that a Point Distribution Model can be used to model constraints on the configuration of features and that the model parameters define a convenient configuration space in which a region representing the set of currently feasible configurations can be maintained. We also introduce the idea of dealing with spatially compact groups of feature hypotheses rather than single hypotheses. We describe two reasoning strategies for dealing with hypothesis groups and feasible configuration regions. These lead to an efficient and exact solution to combinatorially explosive image interpretation problems. We demonstrate the feasibility of the approach by showing results for a system designed to interpret lateral skull radiographs.

1. INTRODUCTION

It is often helpful to decompose a difficult image interpretation problem into a number of simpler problems. Typically, this involves treating a complex object as a collection of sub-parts, or features, which can, to a first approximation, be treated independently. For example, a face may be taken to be composed of eyes, mouth, nose, ears, and so on. Similarly, a mechanical assembly may be treated as a collection of plates, springs, holes, pins, gearwheels and other parts. Generally there are known constraints both on the appearance of individual features and on the configuration of the set of features. If the features are sufficiently distinctive it is straightforward to achieve reliable interpretation by first locating them independently, then verifying that the overall configuration is plausible. If feature detection is less reliable, but there are still a reasonably small number of hypotheses for each feature, it is possible to consider all combinations of feature hypotheses and use the configuration constraints to select those configurations which are globally plausible. There are a number of well-known algorithms for simplifying this combinatorial search when the constraints are rigid [1]. However, when feature location is particularly difficult or configuration constraints are non-rigid the number of combinations which must be considered becomes unworkably large. Unfortunately, many practical applications of vision are like this.

In this paper we describe a system which overcomes some of the difficulties outlined above, by treating feature detection and the application of spatial constraints as co-operating processes. For each feature, a *feature expert* keeps track of all currently valid location hypotheses. A *geometry expert* maintains

the set of all currently feasible configurations of features, represented as a region in a configuration space. If a feature expert is able to discard one of its location hypotheses, some configurations become impossible and the geometry expert can shrink the feasible configuration region. This in turn may exclude some of the location hypotheses for other features. This style of distributed spatial reasoning continues until only a limited number of valid configurations remain. Of these, the configuration most strongly supported by the image evidence is selected as the best interpretation.

We have applied the system to the problem of interpreting lateral radiographs of the skull (cephalograms); the objective is to find a number of bony landmarks which form the basis of a system of measurement used in planning maxillo-facial surgery. The task is challenging and has been the subject of previous research [2,3]. We represent the constraints on the configuration of landmarks with a statistical model, constructed by observing the locations of the landmark features in a training set of cephalograms. To demonstrate the effectiveness of our distributed spatial reasoning strategy we have simulated noisy feature detectors which generate many off-target responses. Given an initial set of feature location hypotheses which combine to give billions of possible configurations, the system is able to find the correct solution by considering only a few tens of configurations.

The idea of using the locations of some features to predict the positions of others has been reported widely [eg 4,5]. The way we deal with non-rigid constraints and our distributed approach to spatial reasoning are, however, new. Several authors describe algorithms for matching under rigid configuration constraints, based on manipulating the bounds of a feasible region in pose space [6,7,8]. Our use of a bounded region in configuration space to represent the set of currently feasible configurations is an extension of this idea.

2. OVERVIEW OF OUR APPROACH

It is probably helpful if we give a brief overview of the elements of our system, before describing the underlying mechanisms in detail. We have already explained that image interpretation is carried out by a collection of co-operating feature and geometry experts. The details of how they interact to arrive at a single solution are covered later; here we outline their roles and responsibilities.

2.1 Feature Experts

Each feature expert deals with a particular target feature; for example, the junction in a cephalogram between the nasal bone and the forehead. It records all current location hypotheses for the feature and maintains a rectangular window which bounds these hypotheses. The hypotheses are generated by running a feature detector over an initial window supplied by the geometry expert. In the experiments we describe below a simple template-matching feature detector has been used. A feature expert may be notified by the geometry expert that its window should be reduced in size. In general this results in

some location hypotheses being excluded. When this happens the window is further reduced in size to just bound the remaining location hypotheses. When a window is resized its new bounds are passed to the geometry expert. A feature expert may also split its window in two; this is described in more detail later.

2.2 Geometry Expert

A geometry expert deals with the configuration of features. The underlying constraints on the spatial relationships between features are modelled using a Point Distribution Model (PDM) [9]. This allows any configuration of points in real space to be represented by a single point in a configuration space. The model is built by analysing a training set of configurations, as a result of which initial bounds can be placed on a feasible region in configuration space. Whenever a feature expert shrinks its bounding window, the geometry expert modifies the feasible region in configuration space to take account of the new constraint. The bounding window for each feature is recalculated and passed to its feature expert. It turns out that calculating the bounding window for a feature in real space, given the feasible region in configuration space, is a linear programming problem.

3. REPRESENTING SPATIAL CONSTRAINTS

A Point Distribution Model (PDM) is a convenient means of representing constraints on the spatial relationships between a set of points [9]. The model can be constructed relatively easily by performing a statistical analysis of the positions of the points in a set of training examples. A PDM can deal either (trivially) with the situation where there are rigid constraints between the points or, more generally, where the configuration is 'flexible'.

A configuration of the points is represented as a vector formed by concatenating all their ordinates $\mathbf{x} = (x_0, y_0, x_1, y_1, \dots, x_{(n-1)}, y_{(n-1)})$. Normally the configurations in the training set are aligned prior to creating this vector representation; in the practical example presented below the cephalograms were always presented approximately upright and centred in the field of view and we used the raw co-ordinates of the landmark features. Given a training set of configuration vectors a Principal Component Analysis is performed. This produces a mean position for each point and a description of the main ways the points tend to move together. New configurations can be generated using the equation

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{P} \cdot \mathbf{b} \quad (1)$$

Where: $\bar{\mathbf{x}}$ represents the mean shape, \mathbf{P} is a $2n \times t$ matrix of the t most significant unit Eigenvectors of \mathbf{S} , \mathbf{S} is the covariance matrix of the training vectors, $\mathbf{b} = (b_1, b_2, \dots, b_t)$ is a vector of shape parameters.

Thus \mathbf{b} space is a configuration space. Each of the parameters b_i controls a mode of variation of the configuration. If each b_i is limited to values less than

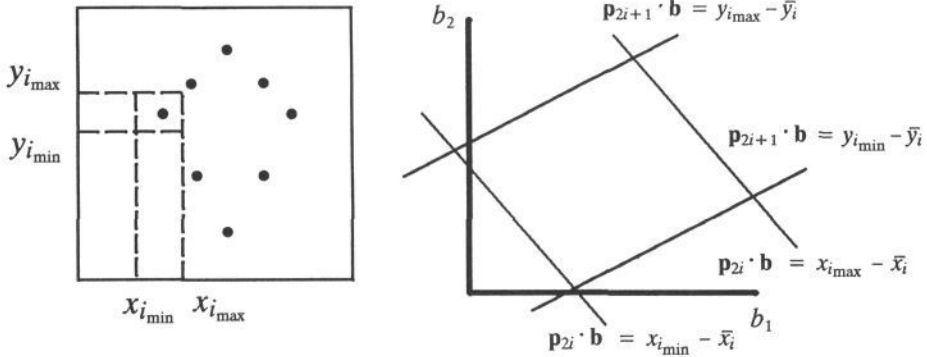


Figure 1. Representing constraints in image space and \mathbf{b} space : when \mathbf{b} space has more than two dimensions the limiting lines become hyperplanes.

\pm a few times its standard deviation measured over the training set then the configurations generated by (1) are similar to those in the training set.

Now consider the relationship between feature windows in real space and the feasible region in \mathbf{b} space. The initial limits on each b_i give a bounded region centred at the origin of configuration space. Suppose we choose one point in real space and apply limits to its position:

$$x_i > x_{i_{\min}} \quad (2)$$

By taking rows of (1) we obtain:

$$x_i = \bar{x}_i + \mathbf{p}_{2i} \cdot \mathbf{b} \quad y_i = \bar{y}_i + \mathbf{p}_{2i+1} \cdot \mathbf{b} \quad (3)$$

Where \mathbf{p}_j is the j^{th} row of \mathbf{P} (a $1 \times t$ matrix).

The inequality (2) can therefore be written:

$$x_{i_{\min}} - \bar{x}_i < \mathbf{p}_{2i} \cdot \mathbf{b}$$

This leads to a limiting hyperplane in \mathbf{b} space, normal to the vector \mathbf{p}_{2i} . All solutions in the half space to one side of this plane are excluded. If the plane intersects with the current feasible region then part of the region can be excluded. When we constrain a feature to lie within a bounding window, four such planes are created in \mathbf{b} space; this is illustrated in figure 1. The geometry expert represents the current feasible region in \mathbf{b} space by the set of hyperplanes which bound it. As new constraints are added they can render existing bounding hyperplanes redundant, in which case they are discarded.

To understand how we can reason with the constraints, consider the case of a model having only two ordinates (one point) and one mode of variation b_0 . This implies that the two ordinates x_0 and x_1 are linearly correlated as shown in figure 2. A limit on b_0 defines a spatial limit for both ordinates. If new

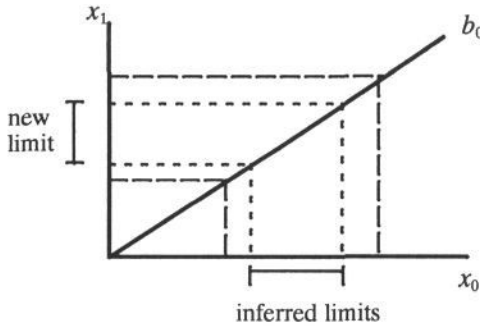


Figure 2. A simple configuration model with two linearly correlated ordinates and thus one mode of variation b_0 . A limit on one ordinate implies a limit on the other via a limit on b_0 .

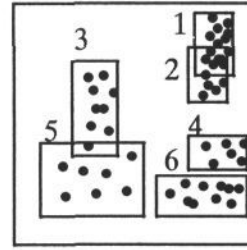


Figure 3. Initial feature windows for the six features used in the cephalogram experiment, each with a set of initial location hypotheses.

limits are found for x_1 then these may be projected through b_0 onto x_0 (dotted lines) giving new limits for x_0 .

4. APPLYING THE CONFIGURATION CONSTRAINTS

We have seen how it is easy to take inequality constraints in real space and represent them in \mathbf{b} space. We have also shown (in Figure 2) how the feasible region in a 1 dimensional \mathbf{b} space can be simply projected back into real space. Unfortunately, when \mathbf{b} space is multidimensional, it is not quite so straightforward to take the current feasible region and calculate the corresponding feature windows in real space. What we need to do is find the maximum and minimum value for each real space ordinate subject to the set of linear constraints in \mathbf{b} space. Since the real space ordinates themselves can be expressed as linear functions of the b_i (see equation 3), this can be formulated as a linear programming problem. The extreme positions of the features correspond to the vertices of the polyhedron bounding the feasible region in \mathbf{b} space and there are well known, efficient procedures for locating them. In our implementation we have used a gradient based linear optimisation method from the NAG library [10].

The computational cost of the linear optimisation method is $O(\min(l^2, t^2))$ where t is the number of modes of variation in the PDM and l is the number of linear constraints – up to 4 for each model point plus two for each PDM mode of variation:

$$l \leq 4n + t$$

5. REASONING WITH SETS OF HYPOTHESES

We are interested in image interpretation problems where many hypotheses are generated for each feature and the number of combinations is too large to search directly: more than 10^{12} in our illustrative example. The basic idea

which we exploit in our solution is that of reasoning with spatially compact sets of location hypotheses rather than individual hypotheses. Configuration constraints can be used to remove whole sets of hypotheses and by keeping the number of sets small the combinatorics can be kept under control. Two reasoning strategies are used. The first involves iteration to achieve consistency between the feature experts and the geometry expert. The second involves splitting the sets of hypotheses handled by the feature experts, creating alternative 'worlds' in which consistent solutions can be sought.

5.1 Finding Consistent Hypothesis Sets

The first step involved in interpreting an image is to establish bounds on the locations of each of the features, based solely on the configuration constraints held by the geometry expert. Sometimes all features may be capable of appearing anywhere in the image, but often the fact that the structures of interest are known to be completely contained within the image, perhaps at an approximately known orientation, allows the geometry expert to initialise the feature windows with loose bounds. Such is the case for the cephalograms we use in our demonstration (see figure 3). Although it is obviously helpful, it is not, however, essential to our method that the initial feature windows occupy less than the whole image.

Once the feature windows have been initialised each feature expert can search its window for its target feature. In general feature detection is unreliable and many off-target responses have to be accepted in order to guarantee that the correct location is included in the set of location hypotheses. It is not necessarily the most efficient strategy for all feature experts to apply feature detection at this early stage but this is of secondary importance; we intend to consider alternatives in due course. Once a feature expert has found a set of feature hypotheses it can adjust its window; in general the rectangle containing all the hypothesised feature locations will be smaller than the region over which they were sought. When a feature expert reduces its window it notifies the geometry expert. Given new, tighter bounds on the location of one feature, the geometry expert can supply tighter bounds to each of the other feature experts. In general, some feature location hypotheses will now fall outside the feature windows for the experts to which they belong. When this happens a feature expert can reduce its window to the point where it just contains the remaining valid hypotheses. Such a reduction can again be passed to the geometry expert, leading to further shrinking of all feature windows. More formally the algorithm is as follows:

1. Initialise feature windows using configuration constraints.
2. Apply appropriate feature detector to each feature window.
3. For each feature window find the reduced window which contains all the detected features.
4. Use these additional constraints together with a configuration model to predict new feature windows.
5. Repeat from step 3 until no further change results.

Where relatively few feature location hypotheses are initially generated, this algorithm can converge to a single valid configuration. In general, however, it will reach a stable state in which at least some feature hypotheses sets have not converged to a single candidate. In this case a second strategy is applied.

5.2 Building alternative worlds

Given a stable, self consistent set of feature groups, further refinement of the locations of the features can be achieved by splitting one of the groups in two. We assume that only one example of each feature is allowed, so the two sub-groups form the basis for alternative solutions. We create two 'worlds' each containing a feature expert carrying some of the hypotheses from the split group together with copies of the geometry expert and all the other feature experts in their current state. Each of these worlds can be iterated as described above; three outcomes are possible: the world converges to a single configuration, the geometry expert finds that no consistent solution can be built from the feature hypotheses in the world, or the world stabilises to a self-consistent state in which more than one solution is still possible. The actions taken are, respectively, to save the solution, stop considering the world, or generate two new sub-worlds. Thus a binary tree of self-consistent worlds is expanded until each leaf node is either 'dead' or a solution. Where more than one consistent solution is found, that with the strongest evidential support is selected; if probabilities can be attached to the feature hypotheses the set with the highest posterior probability should be selected. More formally the algorithm is as follows:

1. Choose a feature expert F and split its window in two, creating two new experts F_1 and F_2 .
2. Construct two worlds based on F_1 and F_2 with copies of all other experts in their current state.
3. Iterate both worlds until they become solved, stable or dead.
4. For any stable world which remains, repeat from 1.
5. For any solved worlds evaluate the evidence supporting the solution.
6. Select the solution with the most support.

This procedure is guaranteed to find any legal solutions which exist. The probability of selecting, from amongst these, the correct solution will depend on the reliability of the measures of evidential support used by the feature experts – it is possible in a noisy system for an incorrect solution to appear more likely than the correct one. It is difficult to generalise about the number of steps needed to reach a solution since this depends on the number of feature location hypotheses generated, their spatial disposition, and above all the strength of the configuration constraints.

6. THE CEPHALOMETRIC APPLICATION

We have tested the methods described above by applying them to a difficult image interpretation problem. Cephalograms are lateral radiographs of the skull taken under controlled conditions. They are used in planning surgery to correct inherited or developmental malformations of the bony structures. The

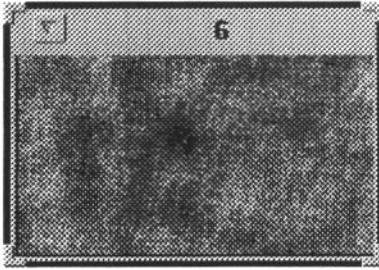


Figure 4: a – The feature coefficient map for one of the initial feature windows..

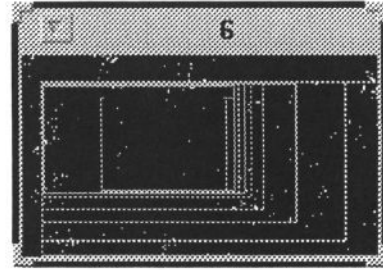


Figure 4: b – Feature location hypotheses sets considered at each iteration.

procedure requires that the positions of a set of bony landmarks are located so that distances, angles etc can be measured. The appearance of each landmark is relatively distinctive, although there are often other structures which can easily be confused. The spatial configuration of the landmarks varies from individual to individual but only within limits.

We built a PDM model of the landmark configurations from a set of 28 cephalograms on each of which six landmarks had been placed by an expert. 99% of the variation in the training set was explained by the 8 modes of variation we used to represent the configuration constraints in a geometry expert. The initial feature windows generated by the model are shown in figure .

In order to concentrate on the spatial reasoning aspects of the problem and have control over the difficulty of interpretation we generated images with the geometry of real cephalograms but with synthetic features. These hybrid images were generated as follows. First a real cephalogram was selected and the positions of its landmark features noted. A uniform image with grey-level g_{mean} was created. For each of six landmarks a 16x16 pixel grey-level template with mean grey-level g_{mean} was generated randomly and placed at the location of the corresponding feature in the real cephalogram. Gaussian noise was added to the image at an amplitude which could be varied. Feature experts used cross correlation with the known landmark templates to detect features. A threshold was applied to create location hypotheses; the threshold level was chosen to ensure that the real feature location was almost always amongst those at which a potential feature was detected.

The system has been implemented in C++ on a Sun SPARC 10. Our ultimate aim is to develop a multi-agent implementation in which control is fully distributed. With this in mind we have simulated messaging interaction between feature and geometry experts which is directly compatible with MAPS, a multi-agent programming system [11].

7. RESULTS

We show results for a deliberately difficult hybrid image with a standard deviation of 2. Figure 4a shows the map of correlation values over the initial feature window for one of the six features used in the experiment. The correct

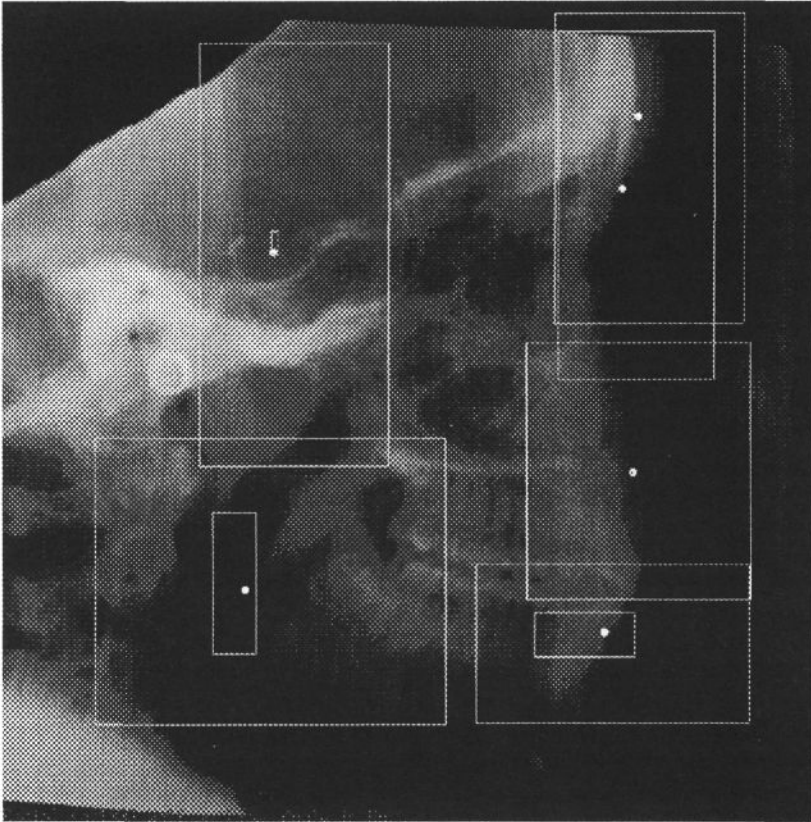


Figure 5. The cephalogram image with feature windows for all six features overlaid.

feature location is near the centre of the window and appears as a bright spot surrounded by a region of lower valued (darker) responses. The hypotheses generated by thresholding the correlation map are shown in figure 4b. The number of feature hypotheses for the six initial windows varied between 4 and 714; the total number of landmark feature configurations possible following feature detection was around 1.5×10^{12} .

Figure 5 shows the situation after 8 window splits; to reach this stage the geometry expert was used to recalculate the feature windows 57 times. Correct solutions have been found for features 1, 2 and 4 and the number of remaining hypotheses for the other features have been greatly reduced; the total number of feature configurations now possible is around 1000. Thus the combinatorics have been reduced by nine orders of magnitude without any significant risk that the correct solution has been discarded. A further 3 splits and 7 calls to the geometry expert were required to reach convergence to the correct solution.

8.0 CONCLUSIONS

We have demonstrated the feasibility of a new method of combining local and global spatial reasoning. Our approach deals effectively with the combinatorial explosion which can arise when many hypotheses are generated for each feature, and allows flexible constraints on the configuration of features to be modelled. We have shown the system being applied to a difficult image interpretation problem based on a medical application. The methods are, however, quite general – the assumptions which have been made are applicable to many practical applications of machine vision. Although we have described image interpretation methods based on the use of PDM models previously the way they are used here is quite new.

There are two main areas in which we envisage new work in the immediate future. First, the criteria used for deciding which feature window to split, when new worlds are needed, require further attention; at present our approach is very ad hoc, though there is no evidence to suggest that the results are sensitive to the selection sequence. More fundamentally, we intend to develop a true multi-agent implementation in which control is fully distributed.

REFERENCES

- [1] W.E.L.Grimson, 1990, *Object recognition by computer: the role of geometric constraints*, MIT Press, Cambridge MA, USA.
- [2] D.N.Davis and C.J.Taylor, 1991, *A blackboard architecture for automating cephalogram analysis*, *Medical Informatics*, Vol 16, pp137–149.
- [3] A.D. Levy–Mandell, A.N. Vestanopoulos and E.D. Filleray, 1986, *An expert system for land-marking of cephalograms*, Proc. of 6th International Workshop on Expert Systems and Applications, pp337–356.
- [4] C. Bolles and A. Cain, 1982, *Recognizing partially visible objects: the local–feature–focus method*, Int. Jour. of Robotic Research Vol 1 No 3, 1982., pp57–82
- [5] N. Ayache and O.D.Faugeras, 1986, *HYPHER: A new approach for the recognition and positioning of two–dimensional objects*, IEEE Trans. PAMI Vol 8 No 1, pp44–54
- [6] T.M. Breuel, 1990, *An efficient correspondence based algorithm for 2D and 3D model based recognition*, MIT A.I. Lab. Memo 1259, Dept. of Brain and Cognitive Sciences / AI Lab. MIT, Cambridge MA, USA.
- [7] T.A. Cass, 1990, *Feature matching for object localisation in the presence of uncertainty*, IEEE 3rd Int. Conf. of Computer Vision, Osaka Japan, pp360–364.
- [8] H.S. Baird, 1985, *Model based image matching using location*, MIT press Cambridge MA, USA.
- [9] T.F.Cootes, C.J.Taylor, D.H.Cooper and J.Graham, 1992, *Training Models of shape from sets of examples*, Proc. BVMC 92 Leeds UK, Eds D.Hogg and R.Boyle, pp9–18, Springer–Verlag.
- [10] P.E.Gill, W.Murray and M.H. Wright, 1981, *Practical Optimisation*, Academic Press, London
- [11] O. Baujard, S. Pesty & C. Garbay. 1991. *A Programming Environment for Distributed Vision System Design*. 6th Int. Conf. on Image Analysis and Processing. Como, Italy.