Chromosome Classification in a General Purpose Frame-based Interpretation System

David Pycock and Christopher J Taylor

Wolfson Image Analysis Unit
Department of Medical Biophysics
University of Manchester
Oxford Road
MANCHESTER M13 9PT

This paper describes work on the implementation of a Chromosome Classification system in which the representation of structural descriptions, generic control strategies and domain specific knowledge is made explicit. A "Best-First" strategy is used to make an initial interpretation of the data which may then be refined using domain specific knowledge.

The work reported here forms part of a larger project to develop 'Techniques for User-Programmable Image Processing' (TUPIP). The objective in this project is to develop a system in which image interpretation tasks can be specified without the need for a detailed understanding of programming techniques and interpretation strategies. The aim is to demonstrate how a system able to operate from a declarative task description could be constructed. As part of this project we are investigating the use of knowledge based techniques to generate a declarative task specification.

INTRODUCTION

In this paper we describe the application of knowledge based techniques in the interpretation of complex data and demonstrate a model based scheme for identifying optimal model matches. We are using a commercial frame-based AI development system to build generic object description and interpretation modules to which domain specific knowledge is added in independent modules. A general discussion of this approach and the issues involved is given elsewhere[5]. Therefore only a brief summary of the principal issues is given in the system overview section.

Most previous work in pattern recognition in general and chromosome classification in particular has focussed on the development of procedural approaches[2,4]. The success of such approaches is limited by the ingenuity and foresight of system developers in anticipating exceptional situations. The promise of knowledge based approaches lays with their potential to enable alternative interpretation strategies to be defined independently of one another and of any particular task. This is achieved by making task requirements and interpretation strategies explicit as separate components of the task specification.

This work is supported by the SERC. The work was started under the Alvey project, MMI-093: 'Techniques for User-Programmable Image Processing' and has been continued under an Advanced Fellowship Award.

Chromosome Classification

During cell division the genetic material of cells is arranged in a set of chromosomes. When suitably processed and viewed down a microscope these chromosomes appear as dark ribbon like objects with a constriction at a characteristic point along their axis. The position of this constriction, measured as a proportion of the length of the chromosome is known as the centromeric index. The size of the chromosome its centromeric index and pattern of stain uptake enables characteristic groupings to be identified and a Karyotype formed as shown in Fig 1. Such karyotype's are used in medical diagnosis to identify genetic abnormalities. It is the recognition of these characteristic groupings which we are seeking to mimic. This is a complex task of global optimisation in which attention must be paid to recognising both the similarity between paired chromosomes and the distinguishing features of characteristic variant forms. It is the complexity which this brings to the description of biological systems such as that considered here which distinguishes this work from other

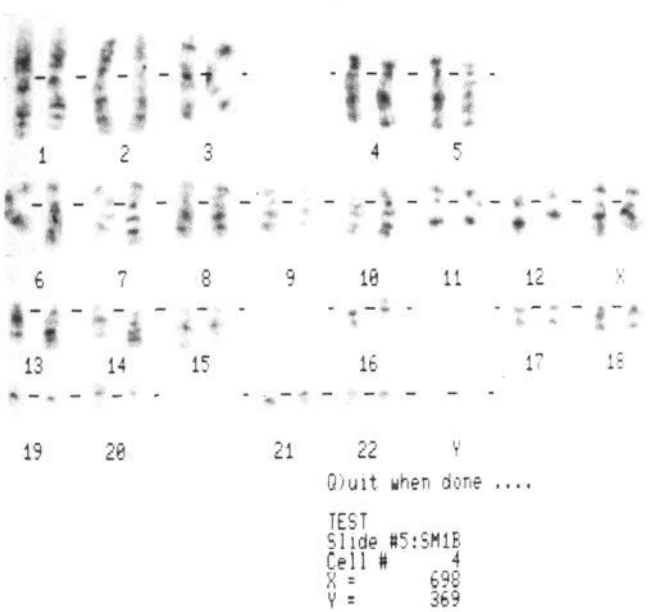

Fig 1. A Chromosome Karyotype

AVC 1989 doi:10.5244/C.3.45

work being carried out in our laboratories on the application of a user programmable vision system to the inspection of complex mechanical assemblies[6].

We have chosen to base our work on Chromosome classification because of our previous experience in developing procedural methods for Chromosome Analysis[2,4] and because it is a well defined and complex classification task for which traditional methods have met with limited success.

Most previous work on the application of knowledge based systems to Chromosome Analysis has sought to address the whole process[3,7]. We believe that it is better to address the many fundamental issues of representation and control from the more constrained aspect of classification before seeking to develop a larger scale scheme for the complete task.

SYSTEM DESCRIPTION

Overview

In previous work under the TUPIP project we investigated the development of an object based image interpretation environment[1]. As a result of that work we concluded that the object paradigm was useful for creating systems which could be described in a tree like structure but that it was not suitable for describing more complex structures such as those required to describe a karyotype as an assembly of chromosomes. This problem arises because for example a karyotype can neither be described as a specialisation nor a generalisation of a chromosome. Rather a karyotype should be represented as an assembly of chromosome entries with both the karyotype and the chromosome described as a specialisation of a common abstract component. Another problem encountered in our previous work concerned the representation of control requirements. The Petri Net model used enabled the specification of control to be made explicit but failed to reduce dependency between modules to an acceptable level. As a result the association between functional methods and objects was in certain cases arbitrary.

The representation used in the work described here overcomes both these limitations. It supports both the construction of hierarchical object based descriptions and the construction of composite structures to describe the assembly of sub-parts. The basic structural unit of the system is the frame or schema. In our current work execution is controlled through a production system. This enables modes of interaction between modules to be separately specified and avoids arbitrary associations between the representation of control and descriptive structures.

Control Strategy

The control of interpretation in this system is by a combination of goal driven backward chaining, data driven forward chaining and model guided strategies. Backward chaining is used to identify achievable sub-tasks by tracing along both the *has-part* and *inheritance* relations. This leads to the generation of goals for the instantiation of each element along the chain until a target component which can be instantiated is identified. Goals are then generated to obtain values for the parameters of the target component. Basic components such as *blob* or *ribbon* features are instantiated first and then used as a basis for the instantiation of higher level *chromosome* and *karyotype* components.

Model based control operates to identify correspondences between empirical data and predefined models. When a complete set of parameters have been obtained for a component an appropriate matching model is identified. In the case of basic components such as *blob* or *ribbon* features this simply involves testing that their size and shape are for example within a predefined range. In the case of components with variant forms such as chromosome models, a bayesian metric is used to identify the most closely matching model.

Finally data driven control is used to update the closeness of match of structures and to update displays.

Model Matching

Closeness of match is assessed by computing the normalised distance in parameter space between data values and model parameters. The distance measures used are normalised with respect to the standard deviation of the model parameters. A match is identified by first identifying the model which most closely matches a candidate. The closeness of match of each candidate to that model is then assessed and the model which most closely matches the candidate chosen. Once a pair-wise optimal match has been identified neither the data nor the model are re-evaluated for matching. This non-exhaustive strategy has been adopted to speed up the identification of initial matches and is based upon the premise that an exhaustive search is unlikely to significantly improve matching. Also the use of a filter to avoid the consideration of very weak matches has been demonstrated to significantly improve performance.

Structural Description

There are two specialisation hierarchies in the system. One is a model hierarchy which is used to describe the distinguishing or prototypical characteristics of structural components. For example the prototypical properties of a *blob* might be its size, compactness and grey level appearance, described in terms of either band limits or the mean and standard deviation of typical values. A *ribbon*, which is a specialised form of a *blob*, inherits the properties of the *blob* and is also described by additional properties relating to its elongation and axial properties. The

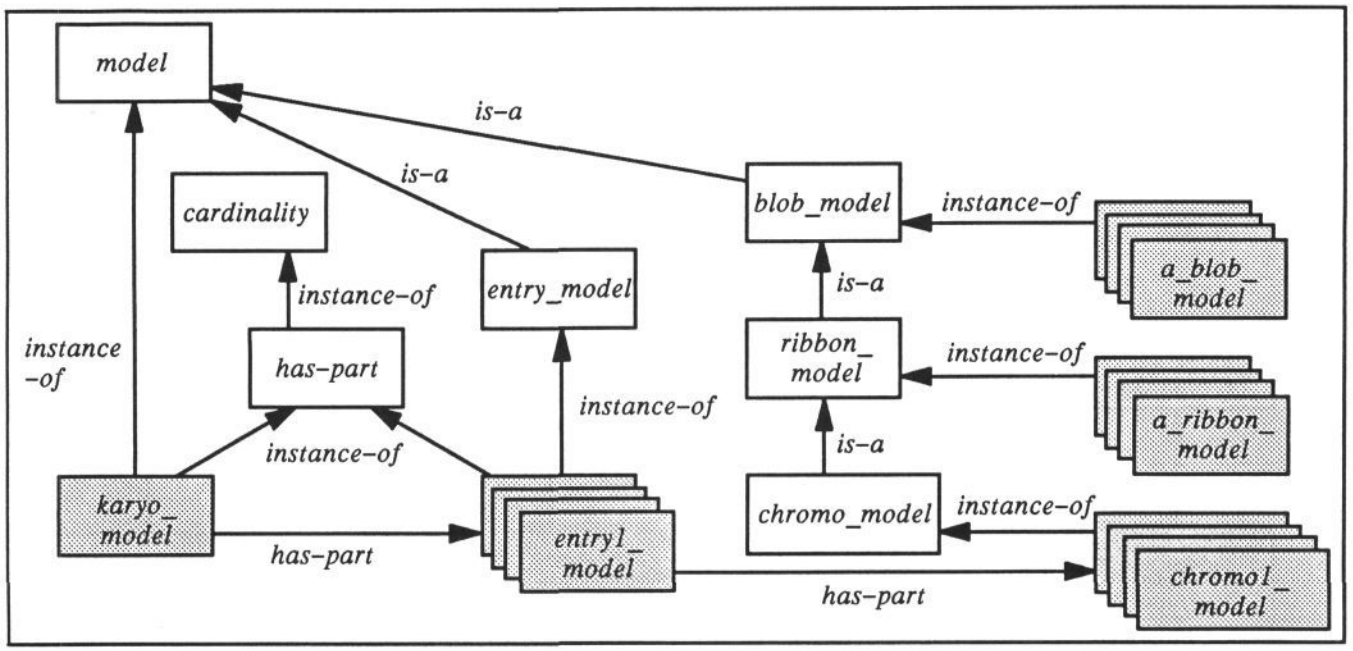

Fig 2. Model description hierarchy.

relationship between typical model components is illustrated in Fig. 2. The model system is equivalent to what in cognitive processes terminology is referred to as medium term memory.

Another hierarchy is used to describe components such as *blobs*, *ribbons* and *chromosomes* which have been identified. The structure of this hierarchy reflects that of the model hierarchy, as shown in Fig 3 and is in effect a mapping between the model representation and the empirical data. This structural description is equivalent to what in cognitive process terminology is described as short term memory.

Relationships

Both reference links and inheritance relations are used in the system. The *has-part* relation is an example of a specially defined inheritance relation used to provide cardinality as a property of the *part - sub-part* relationship.

Two examples of reference links are the *model-of* and *model-of-kind* relations. These reference links are used to associate model descriptions with their instances and with

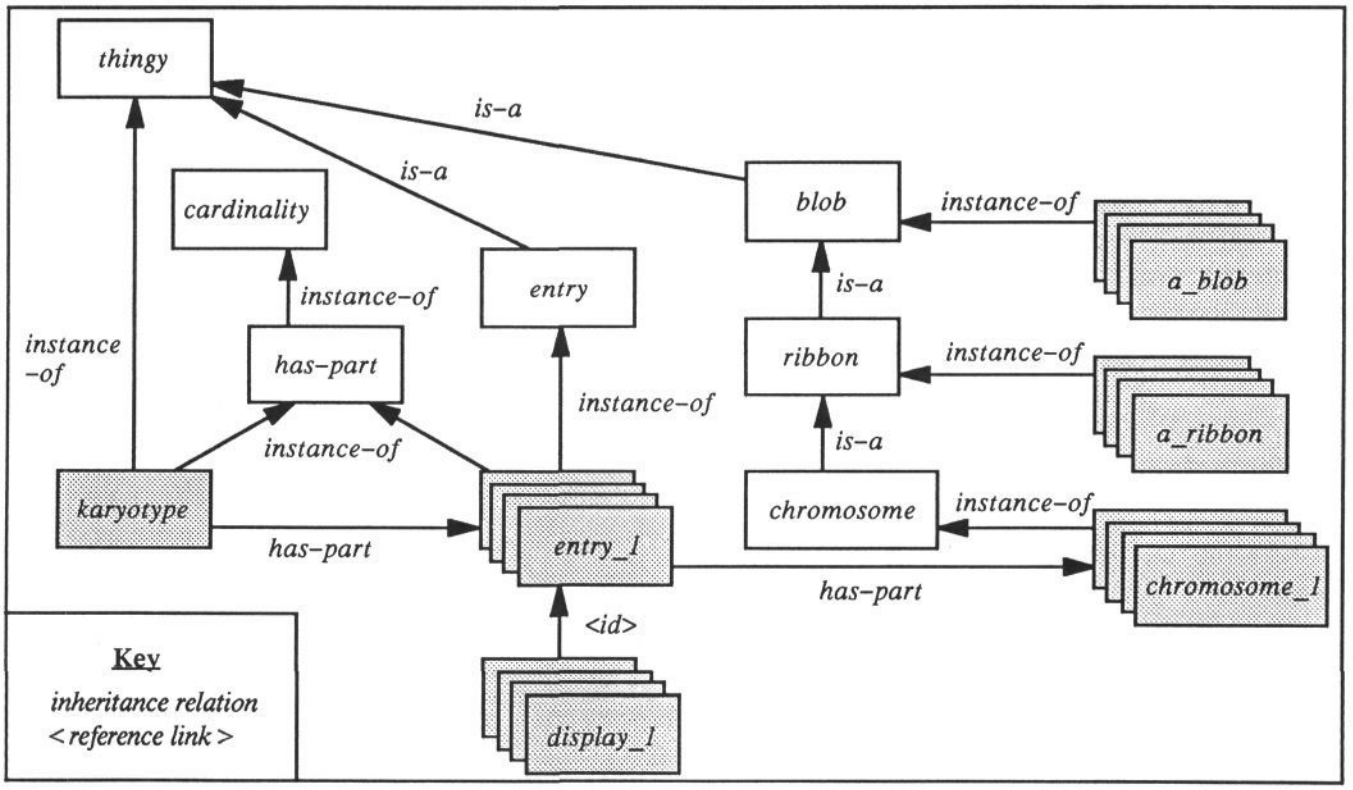

Fig 3. Component description hierarchy.

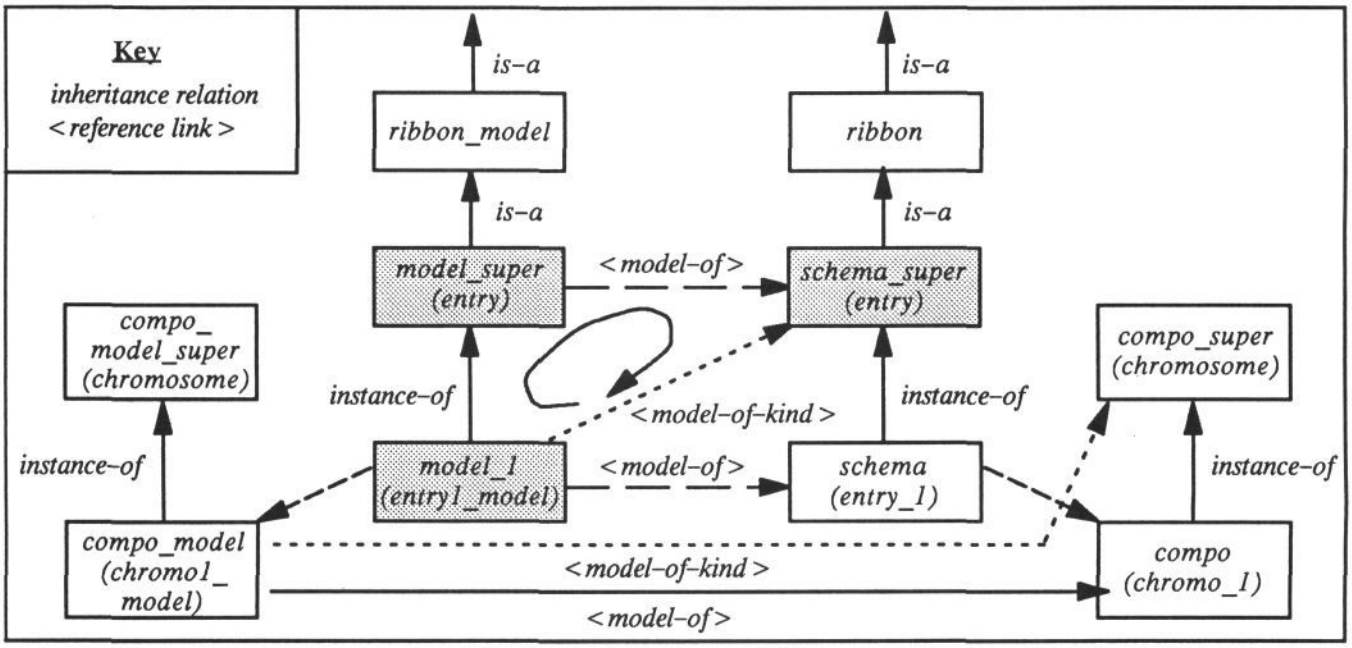

Fig 4. Identification of variant forms of model.

their generic classes respectively. The *model-of-kind* link enables the creation of instances of a particular class to be co-ordinated. This is important where the instances of a class are described by a group of variant models as illustrated in Fig 4. In Fig. 4 the relationships between the three shaded boxes enable all variants of the *entry* class to be identified so that an instance of *entry1* can be created.

Model Description

Most of the knowledge used in the current system is generic with only a little additional domain specific knowledge. Clearly the task specification is domain specific. In this case the task definition includes a definition of how the karyotype is displayed, the number of entries in the karyotype and the number of chromosomes in each entry. The model also defines values for the mean and standard deviation of the size and centromeric index for each chromosome group. There are also a set of priority values associated with each variant model to enable the more easily matched models to be identified and tested first. All the model parameters are programmed as static values but could easily be computed in a training phase from an observation of experimental data.

Implementation

This work has been implemented using ART 3.1 on a SUN 3/160 workstation. The rule base currently consists of 5 display rules and 30 rules for general control. There are currently no application specific rules apart from those used to display the results. However it is expected that a small number of domain specific rules will be needed to recognise when an acceptable classification has been found or to initiate perturbations of the initial classification in seeking to identify a more optimal classification. Such rules would for instance be used to spot the translocation of genetic material between chromosomes.

RESULTS

With a reduced subset of chromosome data designed to test the rule base an error free classification is consistently achieved. The squared bayesian distances used to select model matches are shown for a typical case in Table 1.

Data[1]: Size		46.0	36.0	36.0	33.0	28.0
Cent.		0.475	0.40	0.44	0.275	0.35
Model[2]: Size	Cent.					
41.3 ±3.0	0.50 ±0.16	**2.6**	16.4	33.1	291.8	63.9
39.7 ±2.6	0.39 ±0.02	28.1	**2.3**	11.8	122.6	17.7
32.9 ±2.3	0.46 ±0.02	9.4	7.7	**2.9**	206.0	20.4
30.5 ±2.4	0.27 ±0.01	171.8	36.6	94.8	**1.3**	11.9
24.2 ±2.9	0.35 ±0.05	85.7	23.9	38.1	45.5	**1.7**

Table 1. Normalised distance of chromosome data[1] from model description[2] in parameter space using size and centromeric index (Cent.). The model parameters of mean value and standard deviation and were obtained using 100 cases for each model.

CONCLUSSION

This work has demonstrated the use of knowledge based techniques for chromosome classification and illustrates how inference strategies may be built incrementally using generic mechanisms. The results illustrate that performance similar to that obtained using procedural techniques can be achieved. Clearly the main disadvantage is the time taken to make the classification. This is mainly a characteristic of the development environment used. The techniques employed could easily be transformed into a more efficiently executable form once the principles of design have been identified. Current work is directed towards the development of an interface to image processing modules for data acquisition.

REFERENCES

1. **Bryson, N., Cooper, D.H., Graham, J., Pycock. D., Woods, P.W.,** "DEMOB: An Object-oriented Application Generator for Image Processing" *Proceedings 4th Alvey vision conference,* Manchester, September (1988) pp37–43

2. **Graham, J.,** "Automation of Routine Clinical Chromosome Analysis I Karyotyping by Machine" *Analyt. Quant. Cytol. Histol.,* Vol. 9 (1987) pp383–390.

3. **Piper, J., Baldock, R., Towers, S., Rutovitz, D.,** "A Knowledge-based Chromosome Analysis System" *unpublished report.*

4. **Tso, M.K.S., Graham, J.,** "The Transportation Algorithm as an aid to Chromosome Classification" *Patt Recog. Lett.,* Vol. 1 (1983) pp489–496.

5. **Woods, P.W., Pycock, D., Taylor, C.J.,** "A frame work Sfor modelling and executing visual tasks" *Image and Vision Comput.* Vol. 7 (1984) pp102

6 **Woods, P.W., Taylor, C.J.,** "Application of a User-Programmable Vision System to Inspection of Complex Assemblies" *Proceedings 5th Alvey vision conference*, Reading, September 1989.

7. **Wu Q, Suetens P, Oosterlinck A,** "Toward an expert system for chromosome analysis", *Knowledge Based Systems* Vol. 1, (1987) p43–52.