



Knowledge-based method for segmentation and analysis of lung boundaries in chest X-ray images

Matthew S. Brown^{a, 1,*}, Laurence S. Wilson^b, Bruce D. Doust^c, Robert W. Gill^b, Changming Sun^d

^a*Department of Radiological Sciences, School of Medicine, University of California, Los Angeles, CA, USA*

^b*CSIRO Telecommunications and Industrial Physics, Australia*

^c*St. Vincents Hospital, Sydney, Australia*

^d*CSIRO Mathematical and Information Sciences, Australia*

Received 15 June 1998; received in revised form 17 September 1998; accepted 17 September 1998

Abstract

We present a knowledge-based approach to segmentation and analysis of the lung boundaries in chest X-rays. Image edges are matched to an anatomical model of the lung boundary using parametric features. A modular system architecture was developed which incorporates the model, image processing routines, an inference engine and a blackboard. Edges associated with the lung boundary are automatically identified and abnormal features are reported. In preliminary testing on 14 images for a set of 18 detectable abnormalities, the system showed a sensitivity of 88% and a specificity of 95% when compared with assessment by an experienced radiologist. © 1999 Elsevier Science Ltd. All rights reserved.

Keywords: Knowledge-based segmentation; Chest X-ray; Blackboard; Anatomical model; Edge detection

1. Introduction

Identification of the lung boundaries in chest radiographs is a necessary step for detecting abnormalities such as interstitial disease [1,2], pneumothorax [3], cardiomegaly [4] and pulmonary nodules [5,6].

The aim of this work is to develop an experimental system which demonstrates a knowledge-based approach to segmentation and analysis of the lung boundaries in chest X-ray images.

A clear distinction is made between “high-level” and “low-level” processing. We define low-level processing as operating on raw image data. In such algorithms, the input is considered simply as an array of pixel values. We define high-level processing as operating on data which are represented in a symbolic, knowledge-based domain.

If high-level analysis is to be applied to image data, then a high-level representation must be derived. In our approach, segmentation involves matching low-level image objects to high-level objects described in a mathematical model of the

relevant anatomy (anatomical model). To compare low- and high-level objects, a common, intermediate representation is required, for example using parametric features.

Numerous systems were reported that use of anatomical knowledge, in the form of constraints on features such as expected size, shape, texture and relative positions of structures, to perform image interpretation [7–20]. A number of systems have dealt specifically with segmenting the lung fields in chest X-rays [5,6,21–27]. Typically segmentation is based around thresholding, edge detection and feature-based pixel classification. Duryea and Boone [21] used a heuristic edge-tracing approach with validation against hand-drawn lung contours. Armato et al. [27] used a combination of gray-level thresholding (both global and local) and contour smoothing. McNitt-Gray et al. [23] developed a method using feature-based classification of pixels into regions such as heart, lung and axilla. These systems are effective and useful but do not provide a high-level representation of the image content. Also, domain knowledge is embedded as heuristics within the segmentation algorithms, making it difficult to reapply or extend to other problems.

Images of abnormal anatomy pose a problem for all automated analysis schemes, particularly cases where expected anatomical structures are missing, or altered to the point where they cannot be detected by the available segmentation routines. Most knowledge-based approaches do not

¹Formerly at CSIRO Telecommunications and Industrial Physics, Australia and School of Computer Science and Engineering, The University of New South Wales, Sydney, Australia.

* Corresponding author. Tel.: + 1-310-267-1820; Fax: + 1-310-206-2967; e-mail: mbrown@endeavour.radsci.ucla.edu

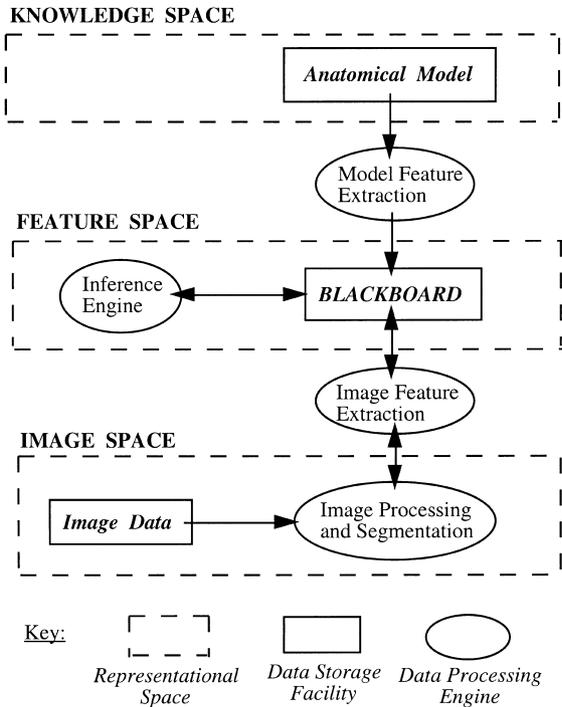


Fig. 1. System architecture and representational spaces.

represent, or reason with, abnormal variations effectively. In some systems, fuzzy sets were used to describe possible ranges of variation in anatomical features, and to quantify levels of confidence during decision-making [13–19].

Fuzzy sets also allow transformation of numerical feature values to symbolic descriptions.

In a knowledge-based approach there is a need for segmentation algorithms, usually a number of them, to interact with the knowledge base. Many computer vision systems, both medical [17,18] and non-medical [28–31], have made use of the blackboard approach to communication and control between the different system components (“knowledge sources”), that contribute to the image interpretation. The blackboard is a data structure that stores the current solution state, and knowledge sources may read from, and write to, the blackboard in an “opportunistic” manner.

If a symbolic description of the image content can be derived, there is the possibility of providing high-level decision support. Many useful non-image based diagnostic and treatment planning expert systems were reported [32–34].

In this paper we present a system that contributes to two main areas of this research:

1. A knowledge-based architecture which combines some aspects of the systems described earlier: an anatomical knowledge base (model), image processing routines, an inference engine for decision-making and control, and a blackboard for communication. Features are used for comparison and matching of image and model objects, and a symbolic description of these features is derived to allow high-level interpretation.
2. Modeling of both normal and abnormal feature values and an inferencing approach which allows segmentation

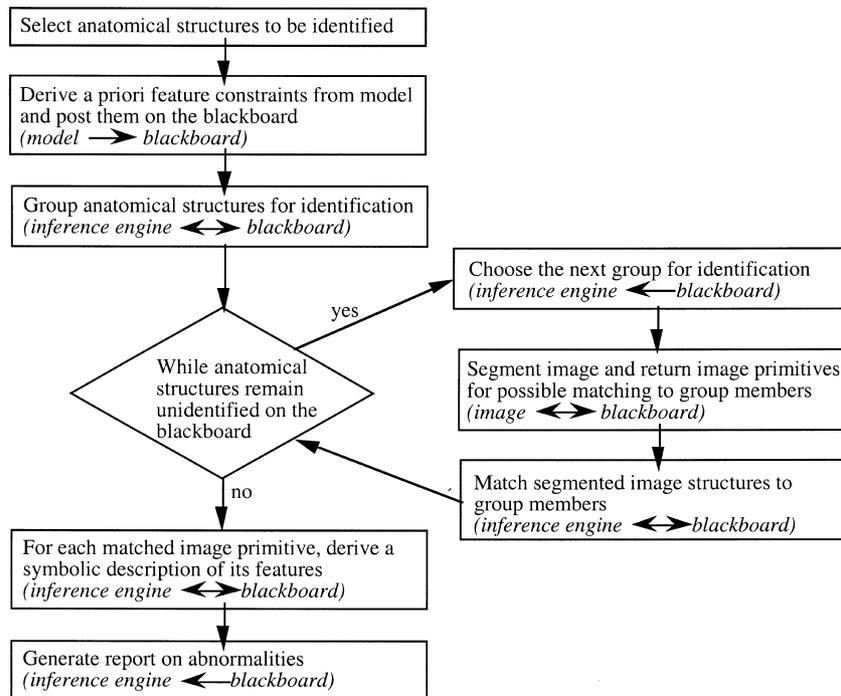


Fig. 2. Overview of system operation, with blackboard interactions in italics.

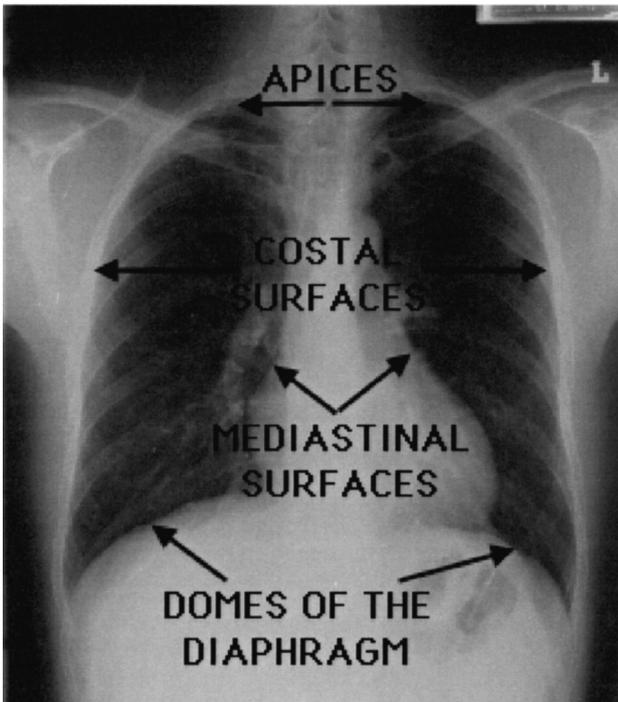


Fig. 3. Anatomically labeled chest radiograph showing lung surfaces.

and matching, even for highly abnormal lung boundaries. There is also a strategy for relaxing constraints in cases where an expected edge cannot be segmented by the low-level routines.

The knowledge-based method and architecture are the focus of this paper, with some preliminary results to show the potential value of such an approach.

2. System overview

The main tasks of the system are to:

- Extract and match image edges to anatomical structures in the model. A common set of numerical feature values are derived from both image and model objects, and comparison and matching occurs in this feature space.
- Derive a symbolic description of the features of matched edges.
- Perform image analysis and generate a report indicating suspected abnormalities. High-level rules translate the symbolic feature descriptions into the natural language report.

The system architecture is based around the blackboard approach to problem solving [35]. Since the components communicate strictly via the blackboard, modularity and independence of internal data representation are possible. Fig. 1 shows the system architecture with the model and image domain representations, and the common feature-space in which they are compared.

In Fig. 2 we outline the control algorithm in terms of systematic interactions with the blackboard. Anatomical structures that are interrelated are grouped for matching, and identified simultaneously. Opportunistic scheduling occurs in the step where the next group of anatomical structures is selected for segmentation (see Section 3.4.1).

3. System elements

3.1. Anatomical model

3.1.1. Organization of the knowledge base

For each anatomical structure to be identified, the system

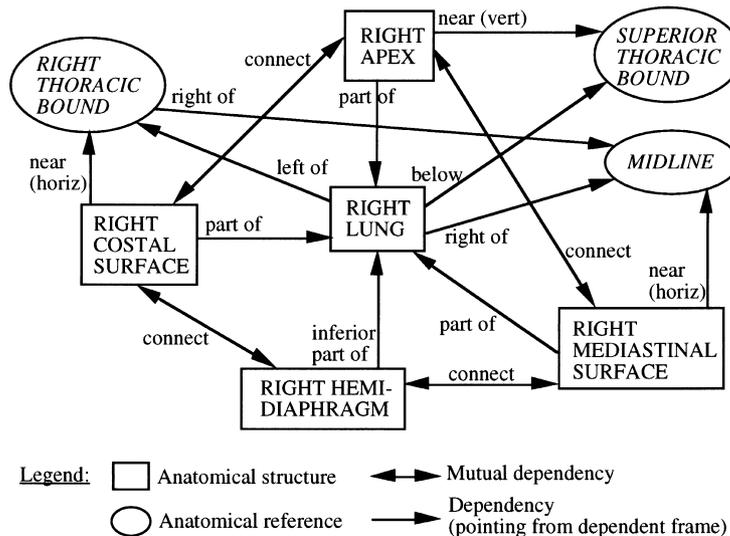


Fig. 4. Semantic network for frames involving the right lung.

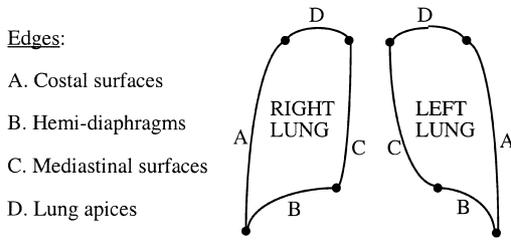


Fig. 5. Edges included in the lung boundary model.

uses the following knowledge: name, shape information, structural relationships to other anatomy and imaging characteristics. The model focuses on the lung boundaries, which are significant because their shape is determined by the surrounding anatomy: the domes of the diaphragm, mediastinal (and cardiac) silhouette, lung apices and costal margins (bounded by the rib cage) are indicated in Fig. 3.

There are a number of ways in which the domain knowledge can be represented. While the various representations may be capable of encoding the same information, there are differences in terms of the ease with which application systems can be built, as well as the extensibility and maintainability of the knowledge base.

A rule-based knowledge representation was used in systems for identifying blood vessels in angiograms [7,12]. Rules are not a natural representation for many of the declarative facts required in our knowledge base, and this limits its effectiveness as described earlier. However, we use a rule-based approach for generating reports on abnormalities (see Section 3.4.4).

Voxel-based models or atlases were used to represent anatomical information, and deformed to fit individual data from CT and PET images [36]. However, there are difficulties in representing possible anatomical variations and voxel classifications are typically derived from a single cadaver. Some methods for segmenting brain MR images have attempted to overcome this by using multiple subjects to create probabilistic spatial distributions of normal anatomy [37,38].

In our system, anatomical knowledge is stored in a declarative model. For each anatomical structure, parametric shape and relational attributes are encapsulated in a frame [39]. The frames have a predefined set of “slots”, corresponding to the attributes, in which parameter values are stored.

Table 1
Relational vocabulary used in the model and associated parameters

Relationship	Parameters
Part of	
Right of	Distance to the right
Left of	Distance to the left
Inferior to	Inferior distance
Superior to	Superior distance
Near (horizontal)	Horizontal distance
Near (vertical)	Vertical distance
Connected to	Distance between endpoints of the edges
Inferior part of	Distance below centroid
Superior part of	Distance above centroid

Relationships between anatomical structures are represented by links between the frames in the model, as in a semantic network [40]. Relational slots store a pointer to the related frame and associated parameters. Frame-based semantic networks were used previously by a number of groups [8,10,15]. Part of our semantic network, dealing with the right lung, is shown in Fig. 4. The complete network consists of 14 frames and 30 relationships.

3.1.2. Parametric description

Here we describe the parameters used to model shape and structural relationships between anatomical structures.

The shape description is edge based. Edges are modeled as curves with expected length, position, orientation at the endpoints, and edge strength as parameters. Lengths are modeled in centimeters, and are later converted to number of pixels for a given image. We specify endpoint orientation relative to an (x, y)-coordinate system, where the x-direction is to the patient’s left, and the y-direction is inferior. Smooth curvature is assumed between the endpoints. We define the edge strength as a percentage of the maximum strength of any edge which is expected in the image. This makes the parameter less dependent on the method used to compute edge strength (e.g. image gradient magnitude). Fig. 5 shows the edges which are included in the model, and Fig. 6 shows the slots in a model frame.

A vocabulary was defined to model relationships between anatomical objects. Table 1 shows the relationships and associated parameters.

3.1.3. Modeling anatomical variability

An important requirement of the knowledge

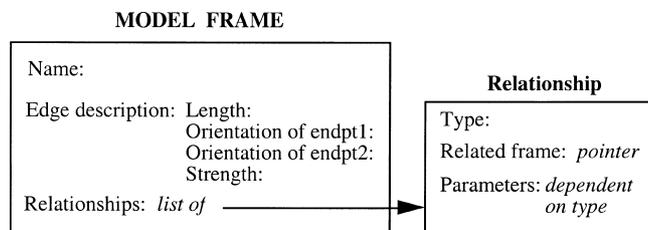


Fig. 6. Model frame structure.

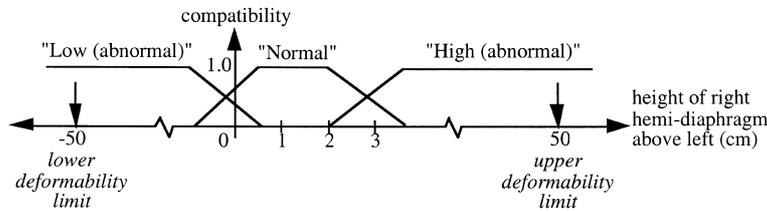


Fig. 7. Modeling anatomical variability using fuzzy sets and deformability limits.

representation is the ability to describe anatomical variations between patients. Each parameter in the model defines a range of numerical values which are considered possible (see Section 3.1.3.1), and a fuzzy mapping between numerical values and linguistic (symbolic) descriptions (see Section 3.1.3.2).

3.1.3.1. Numerical parameter variability The model describes anatomical variations which can occur owing to elastic deformations of the lung. It represents variations such as those caused by structures becoming enlarged (e.g. cardiomegaly), reduced (e.g. collapsed lung) or displaced (e.g. by swelling of an adjacent organ).

For each parameter in the model, upper and lower “deformability limits” specify the possible range of values under natural deformation. This is, and must be, a coarse description because we are modeling possible, rather than normal, variations. If the model was more tightly specified, the system’s ability to recognize unusual anatomy would be compromised. Because crudely specified deformability limits are used, they do not have to be redefined for subjects of different sex or race, and are usable over a wide range of ages.

3.1.3.2. Linguistic imprecision Anatomical descriptions are subject to the imprecision of natural-language. Fuzzy

set theory pertaining to linguistic variables [41] allows representation of this vagueness in mapping between numerical parameter values and linguistic descriptions. For each linguistic expression, a fuzzy set is created which provides compatibility (fuzzy membership) scores for possible parameter values.

Fuzzy sets were used previously to map numerical values to symbolic representations for high-level matching in medical images [13,17,42]. Typically, the same set of linguistic variables is applied to a given feature for all anatomical structures. For example, the linguistic variables, *small*, *medium* and *large* may be used to represent size for all organs. In our approach, we redefine linguistic variables for each anatomical structure in the model because, for example, a *small* lung represents different numerical values from those for a *small* thorax. Further, rather than modeling a structure as being *small* or *large* we believe that a more natural description can be derived relative to *normal*, with structurally unusual variations being modeled as, for example, *abnormally small* or *abnormally large*. Such a representation is based on the assumption that a major part of diagnosis involves recognizing “normal” anatomy, and then excluding it from further attention, and therefore experts are readily able to supply constraints in this form. Here we are using *abnormal* simply to describe the appearance of a particular feature, at this level there is no implication that it is necessarily the result of disease.

Consider as an example, the relationship between the heights of the right and left hemi-diaphragms. Fig. 7 shows an example of fuzzy compatibility functions for the height parameter, as well as deformability limits. The normal compatibility function is relatively narrow, however the deformability limits are a long way apart because the diaphragm is flexible, i.e. highly deformable. Thus, the model indicates that under normal conditions the relative heights of the hemi-diaphragms are quite well-defined, but a wide range of abnormal variation is possible.

For this preliminary work, the fuzzy compatibility functions and deformability were determined empirically with guidance from an expert radiologist. They were checked and refined against measurements made on a training set of 12 chest X-rays. We have found trapezoidal (piecewise linear) compatibility functions to be sufficiently precise for our current application. In a statistical approach, such functions could be generated based on measured feature values from a set of training images. This would require a representative

Table 2

Translation of relational vocabulary to spatial constraints in the image. M = set of pixels in image mask for dependent object; B = set of pixels in related (parent) object; P = relationship parameter range (see Table 1); P_L = lower deformability limit; P_U = upper deformability limit; $B_{xmin}(y)$ = minimum x such that $(x, y) \in B$; $B_{xmax}(y)$ = maximum x such that $(x, y) \in B$; $B_{ymin}(x)$ = minimum y such that $(x, y) \in B$; $B_{ymax}(x)$ = maximum y such that $(x, y) \in B$

Relationship	Meaning in image domain
Right of	$M = \{(x, y) (B_{xmin}(y) - P_U) < x < (B_{xmin}(y) - P_L)\}$
Left of	$M = \{(x, y) (B_{xmax}(y) + P_L) < x < (B_{xmax}(y) + P_U)\}$
Inferior to	$M = \{(x, y) (B_{ymax}(x) + P_L) < y < (B_{ymax}(x) + P_U)\}$
Superior to	$M = \{(x, y) (B_{ymin}(x) - P_U) < y < (B_{ymin}(x) - P_L)\}$
Near (horizontal)	$M = \{(x, y) (B_{xmin}(y) - P_U) < x < (B_{xmax}(y) + P_U)\}$
Near (vertical)	$M = \{(x, y) (B_{ymin}(x) - P_U) < y < (B_{ymax}(x) - P_U)\}$

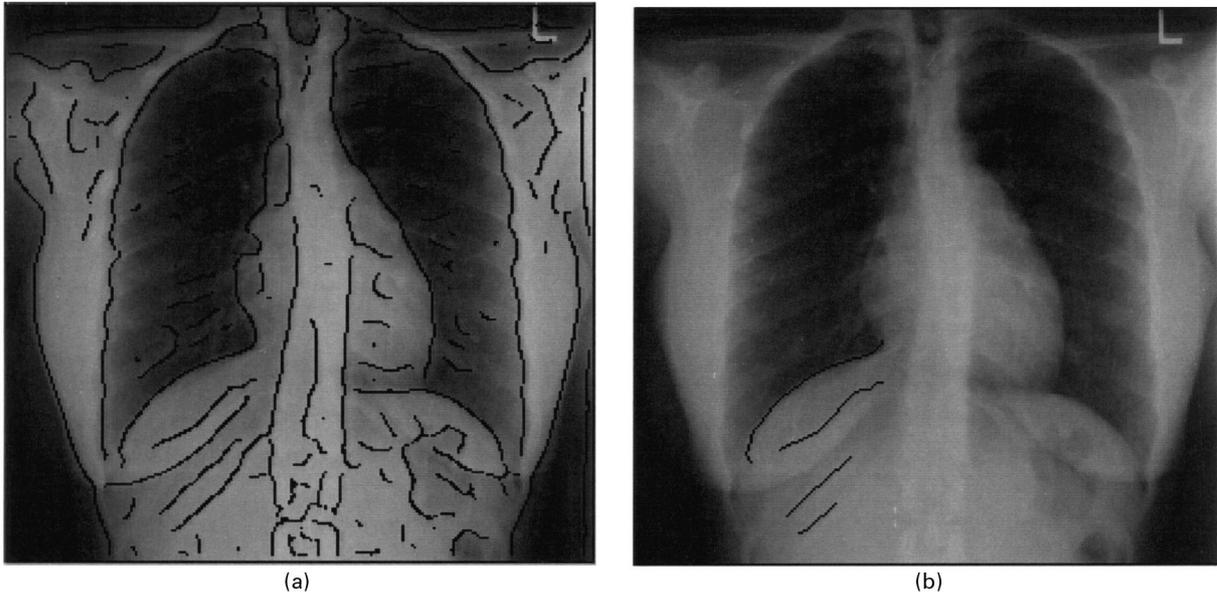


Fig. 8. (a) Canny edge detection with $\sigma = 3$; (b) Candidates generated for the right hemi-diaphragm.

training set which is difficult because of variations with age, sex and race, as well as the many different abnormalities which can occur.

3.2. Image processing

3.2.1. Knowledge guidance

The anatomy which should be present in most medical images can be stated a priori and anatomical knowledge can be used to guide segmentation routines to extract specific structures. Thus, our image processing is primarily model-driven.

The system uses both a priori (model-derived) and a posteriori (image-derived) knowledge to spatially constrain the segmentation of a given anatomical structure. Structural

relationships to previously labeled anatomy are used to refine an image mask which specifies the image region where edge extraction should be performed. Table 2 defines the relational vocabulary in terms of translation to constraints on pixel coordinates (see Section 3.1.2 for coordinate system).

When an anatomical structure is scheduled for segmentation, knowledge-based constraints are read from the blackboard and passed to the image processing engine. The extracted edges are written back to the blackboard as “candidates” for matching to the anatomical structure.

The following constraint parameters are passed to each edge detection algorithm: strength (see Section 3.1.2 for definition), orientation (gradient direction), position (image mask), and length (number of pixels). These

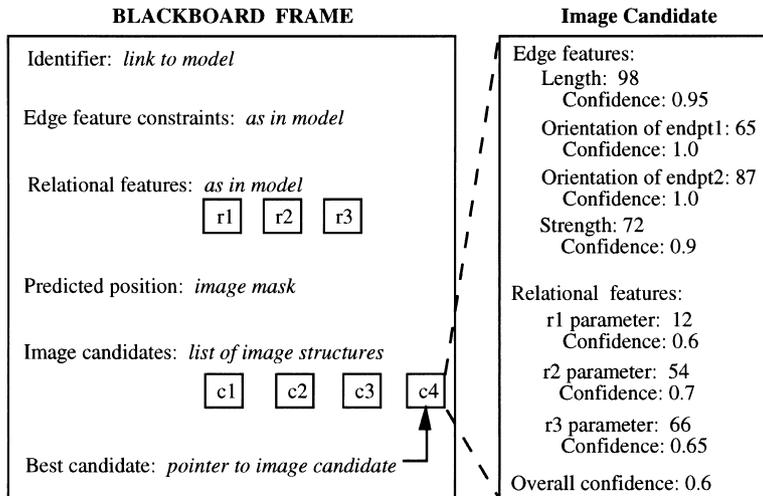


Fig. 9. Structure of a blackboard frame.

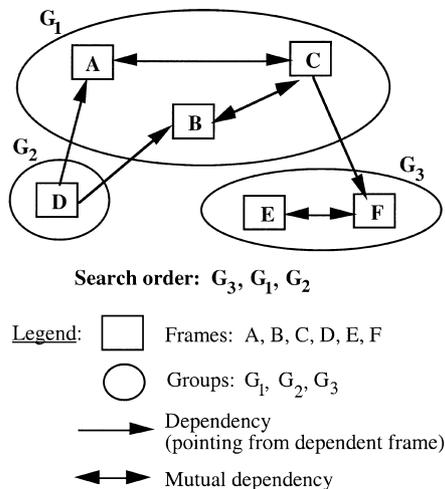


Fig. 10. Example of the grouped search strategy.

constraints are passed as numerical ranges, corresponding to the deformability limits, which are read from the appropriate slots in a blackboard frame.

3.2.2. Edge detection

Canny edge detection is performed, but typically the expected anatomical structures do not produce a single, continuous edge. Therefore the image processing engine applies the Canny operator at multiple resolutions, with standard deviations, $\sigma = 3, 7, 9$ (the exact choice of values is not critical), and uses the following algorithm to generate candidates by linking different combinations of edge fragments:

1. apply gradient magnitude (strength) and phase (orientation) constraints;
2. form edge fragments from continuous sets of points;
3. subdivide (split) edge fragments at points of high curvature; and
4. calculate possible linkages of edge fragments to form candidates: a linkage can occur if the orientation of a line joining the end points of the fragments is within the orientation constraint range, and the overall length constraint is satisfied.

Fig. 8 shows an example of edge detection for the right hemi-diaphragm. There will typically be 5–10 edges (candidates) generated for an anatomical structure.

3.3. Blackboard

The blackboard uses frames to store feature-space information. For each anatomical structure to be identified, a frame is created and posted on the blackboard. Slots exist for storing data derived from both model and image, as well as matching results produced by the inference engine (see Fig. 9).

In mapping the model to feature-space, constraints on feature parameters are derived from the model and written

to blackboard frames. Edges are represented in feature-space using the same parameters as defined in the model (see Section 3.1.2). Relational links are created between the frames as required. Each expected feature value (constraint) is represented by a “confidence function” (definition and derivation given in Section 3.4.2.1) and fuzzy sets taken directly from the model.

3.4. Inference engine

The inference engine performs a number of tasks related to decision-making within the system:

- scheduling the order in which anatomical structures are identified;
- selecting the best edge (candidate) for matching to an anatomical structure;
- deriving symbolic descriptions of feature values; and
- high-level testing for abnormalities and report generation.

3.4.1. Scheduling

Since the frames on the blackboard are relationally constrained (linked), object recognition requires the best combination of candidates to be determined for matching to the set of frames on the blackboard, as opposed to finding a best candidate for individual frames independently.

Given the number of frames and candidates which exist, an exhaustive search of all possible combinations is not feasible. Alternatively, an ordered search could be applied where an image structure is matched to a frame, and then remains fixed and is used to guide the search for other structures. This is much faster, but may produce sub-optimal results.

Our scheduling strategy is a compromise between the exhaustive and ordered searches. It is based on the hierarchy implied by the relationships in the model. Binary relationships define dependencies, both uni-directional and “mutual” (defined later), between frames on the blackboard. Mutually dependent frames are grouped for searching as follows: a frame, B , is a member of group, G_1 if there is another frame, $C \in G_1$, such that B and C are mutually dependent (see Fig. 10). Within a group, an exhaustive search is carried out to find the best combination of candidates (i.e. best group candidate). The match to frames within the group is then fixed and then the next group is processed, giving an ordered search between groups.

The number of a mutual dependencies determines the size of groups and hence the computation time of the exhaustive search. For the purpose of forming a group, a relationship is considered a mutual dependency if two criteria are met:

1. The type of relationship must be defined as a potential mutual dependency in the relational vocabulary. Many relationships could sensibly be defined as either a mutual or uni-directional dependency. For example, we define *connected* to as a potential mutual dependency.

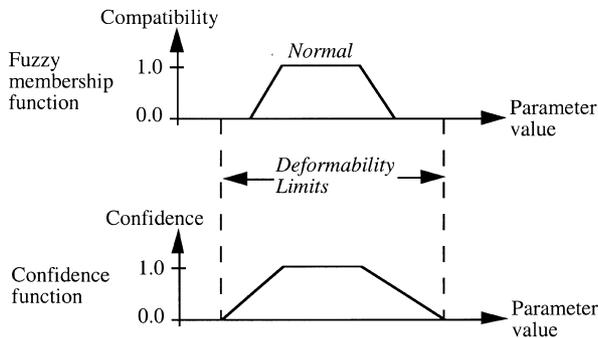


Fig. 11. Initialization of a confidence function using the normal fuzzy set and deformability limits.

Therefore it is appropriate for describing a connection between structures of similar significance (in terms of labeling). However, a new uni-directional dependency would need to be defined for a connection between a major structure and a smaller or more subtle structure. Fig. 4 shows which relationships are defined as potentially mutual.

2. The relationship must have a high degree of “invariance”, since these are most important in object recognition. The amount of variability in a relationship is often determined by the flexibility, or rigidity, of the structures involved. The degree of invariance is quantified, based on the model parameter associated with the relationship, as the ratio of the width (support) of the *normal* compatibility function to the width of the deformability limits. A relationship is considered a mutual dependency if this ratio is greater than 0.5.

Only uni-directional dependencies exist between frames in different groups and the dependencies imply a search order. A uni-directional dependency is “usable”, in terms of applying constraints on a frame, if a match was made to its “parent” frame. The next group to be identified is chosen as the one with the fewest unusable dependencies, so that as many constraints as possible are available to guide the segmentation. This grouped search, illustrated in Fig. 10, is a compromise between speed and an optimal solution.

The formation of the groups is dependent on the relationships stated in the model, and thus has a weakness in that it assumes the relationships will naturally define a hierarchy suitable for searching. The advantage is that the method can be fully automated, allowing the system to compute a schedule for identifying any subset of anatomical structures described in the model.

3.4.2. Matching

The best candidate for matching to a frame is selected on the basis of confidence scores. For each feature associated with a candidate, a confidence is calculated which indicates how well the candidate satisfies the constraint on that feature.

There are four issues to be explored in computing a confidence score for a candidate:

1. generating confidence scores for individual constraints;
2. combining these scores into an overall confidence for the candidate;
3. selecting the best candidate using the confidence scores; and
4. dealing with the case where no suitable candidates can be found.

3.4.2.1. Generating confidence scores: confidence functions For each feature parameter in the model, a confidence function is derived which maps parameter values to confidence scores for matching purpose. It is assumed that each model parameter will have a fuzzy set defining the linguistic value, *normal*, and that it will have a trapezoidal compatibility function. Further, each parameter should have deformability limits defined. The derived confidence function is also trapezoidal, such that it has value 1 wherever the normal compatibility function is maximum, and decreases linearly to 0 at the deformability limits as shown in Fig. 11. This gives an a priori approximation of the range of possible values which a parameter can take under natural deformation. Invariant parameters (with narrow deformability limits) have tightly defined confidence functions, while more variable parameters will have broad confidence functions.

3.4.2.2. Combining confidence scores To generate an overall confidence score for a candidate, a fuzzy logic approach is used, i.e. the minimum of the confidence scores for individual features is taken. This fuzzy logic method has a potential weakness of being sensitive to the most poorly satisfied constraint only. Our inferencing technique addresses this by virtue of the loosely defined confidence functions (based on deformability limits), so that all constraints should be satisfied with high confidence for a valid candidate. Under these conditions fuzzy logic is an acceptable means of obtaining an overall confidence measure. Loose definition of constraints may appear to present a problem in terms of discriminating between structures, however this is countered by three facts:

- Many structural relationships are naturally invariant, so they do not have to be initialized loosely.
- Relational (spatial) constraints impose tight a posteriori constraints in addition to the loosely defined a priori constraints.
- Combination of many constraints for matching allows them to be more loose than if a small number were used.

3.4.2.3. Selecting the best candidate The task is, in fact, to select the best combination of candidates for a group, rather than dealing with individual frames independently. A “group candidate” is formed for each possible

Table 3
Rules for generating reports of anatomical abnormalities (symbolic feature descriptions in *italics*)

Reported anatomical abnormality	Feature-based rule
Small lung volume	Length of costal edge: <i>short (or long)</i>
Over-inflated lung	
Asymmetric lung volumes	
Silhouette sign	Edge: <i>unfound</i>
Unclear interfaces	Edge strength: <i>weak</i>
Raised hemi-diaphragm	Height of the right hemi-diaphragm above the left: <i>low (or high)</i>
Abnormal lateral portion of hemi-diaphragm (query pleural effusion)	Orientation at endpoint of hemi-diaphragm: <i>abnormal</i>
Unclear cardiac silhouette	Length of mediastinal edge: <i>short</i> , or strength of mediastinal edge: <i>unclear</i>
Widening of the upper mediastinum	Distance between upper endpoints of the mediastinal edges: <i>enlarged</i>
Cardiomegaly	High cardiothoracic ratio, i.e. max. width of mediastinal edges/max. width of costal edges
Abnormal lung density	High average gray-level of lung region
Asymmetry of lung density	Abnormal density in single lung

combination of candidates from member frames. The confidence of a group candidate is computed as the minimum confidence of its member candidates, and the best group candidate is selected.

Since our matching strategy uses an ordered search between groups, without backtracking, there may be some concern regarding cascading of matching-errors. The system only allows a “good” match to a frame, otherwise the frame is labeled *unfound* and does not effect the matching of subsequent groups. The system aims for good matches by using multiple constraints and demanding that they all be satisfied (via fuzzy logic).

3.4.2.4. Constraint relaxation The system is able to relax constraints under certain conditions to increase robustness for noisy, incomplete or unexpected data. If an image edge predicted by the model cannot be detected by the available detection algorithms (i.e. the correct edge is not generated), some (incorrect) candidate edges may still be produced. These candidates will typically fail to satisfy at least one constraint, and thus will have zero confidence for matching. In such cases, all group candidates (for the group to which the frame belongs) have zero confidence, even though other frames in the group may have good (correct) candidates. Constraints on the group must be relaxed if a match is to be made.

The inference engine relaxes constraints by adding the *unfound* candidate to one or more frames in the group, thereby creating some new group candidates. The *unfound* candidate has the property that for any constraint in which it is involved (including relational), it yields a confidence score of 1 (i.e. complete confidence). Thus, if a frame is labeled *unfound*, it imposes no constraints on related frames.

The minimum number of frames are labeled *unfound*, such that there is at least one new group candidate with non-zero confidence.

This constraint relaxation is coarse, in the sense that it effects all constraints associated with a particular frame, rather than allowing individual constraints to be relaxed. This is consistent with the assumption that all constraints must be satisfied. If any one of the constraint fails, then a candidate cannot, and should not, be matched to the frame.

3.4.3. Symbolic feature description

Features of recognized anatomical structures are described symbolically using the fuzzy sets defined in the model. For each parameter value, the system generates a linguistic value and a compatibility score indicating how well the description fits the value.

The symbolic representation provides important complementary information to the confidence scores generated during matching. While the confidence scores deal with loosely defined ranges of possibility, the linguistic values give an assessment against more strict criteria for normality.

3.4.4. High-level rules for reporting abnormalities

The system uses high-level rules to generate reports on suspected abnormalities based on the symbolic feature descriptions. Most of the abnormalities we are concerned with, can be detected by analyzing the lung boundary. We describe briefly some examples:

- **Silhouette sign**
An increase in density of the lung (e.g. owing to fluid or calcification) may cause one, or more, of its interfaces to become unclear.
- **Lobar collapse**
Collapse of one of the lobes of the lung can be detected from the image by the reduction in lung volume and raised position of the hemi-diaphragm.
- **Pleural effusion**
Fluid in the pleural space often collects at the base of the lung, giving a rounded, rather than sharp, costophrenic angle (where the costal interface meets the dome of the diaphragm).
- **Cardiomegaly**
An enlarged heart, the size of which can be estimated from the image.

We are also interested in detecting abnormalities in lung density, either asymmetry between the lungs or inhomogeneity within one of the lungs. The rules used to detect specific abnormalities are shown in Table 3. The compatibility scores for the linguistic values are used to calculate a confidence score, again by fuzzy logic, for each abnormality reported via the rules.

3.5. Implementation

The system was implemented on a Silicon Graphics

Table 4

Comparison of system output against a radiologist’s assessment (for limited set of abnormalities): ● reported by both (true-positive); ○ reported by radiologist only (false-negative); × reported by system only (false-positive)

Image identifier	A	B	C	D	E	F	G	H	I	J	K	L	M	N
<i>Lung volume</i>														
Small right													●	
Small left										×		●		●
Asymmetric										×		●	●	●
<i>Lung density</i>														
High right														
High left		○								●		○		
Asymmetric		●								●		●		×
<i>Unclear silhouette</i>														
Right costal														
Right mediastinal				×										
Right hemi-diaphragm				○										
Left costal										●		●		
Left mediastinal		●								●		●		
Left hemi-diaphragm						×		●		●		×		
Cardiac		●		×						●		●	●	●
<i>Raised hemi-diaphragm</i>														
Right													●	
Left										×		●		●
<i>Abnormal lateral hemi-diaphragm</i>														
Right								●						
Left						●	×	○			×	●		●
<i>Cardiomegaly</i>														
								●						

Indigo 2 workstation. Common LISP (with Object System extension) was found to be convenient for creating frames and performing inferencing and control. Therefore, most of the system architecture is coded in LISP, with the image processing routines being implemented in C.

4. Experimental results

4.1. Experiments

Some preliminary experiments were performed to gauge the potential of the approach. A set of 14 images was used for testing, acquired by computed radiography (CR) or from films scanned on a laser digitizer. All images were subsampled to dimensions of approximately 256 × 256 to allow fast image processing. This resolution was adequate given that we are currently only analyzing large-scale anatomy. Total processing time was approximately 2 min on the Indigo 2. The use of higher-resolution images does not effect the speed of matching and analysis, since these processes deal with parametric feature descriptions of

Table 5

Summary of test results, where an abnormal finding is considered a positive test result

	True	False
Negative	208	4
Positive	30	10
Total	238	14

image primitives, however, the low-level computation of the edge images is slower.

The test images were not chosen randomly. They were selected with the aim of having 50% normal and 50% abnormal images such that most of the detectable abnormalities were covered. Five of the images were from our training set and nine were previously unseen test images.

For each set image, an experienced radiologist was asked to identify abnormalities from the set of 18 specified in Table 4. These abnormalities were compared against those reported automatically by the system.

4.2. Results

Table 4 shows the abnormalities which were reported by the radiologist and the system for each image. For the total of 252 tests for abnormalities, the system performance is summarized in Table 5, where an abnormal finding is considered a positive result.

We identify two main factors contributing to the small number of false positive and false negative test results: (a) for a low-contrast edge, the edge detection is often inaccurate; (b) the inference engine produces a confidence score for each reporting decision, and Table 4 includes all non-zero confidences as positive, therefore some reported abnormalities with very low confidences are false positives. Five of the false positives were reported with a confidence less than 0.4.

Fig. 12 shows the system output for two images of normal anatomy, with the detected lung boundary and reports

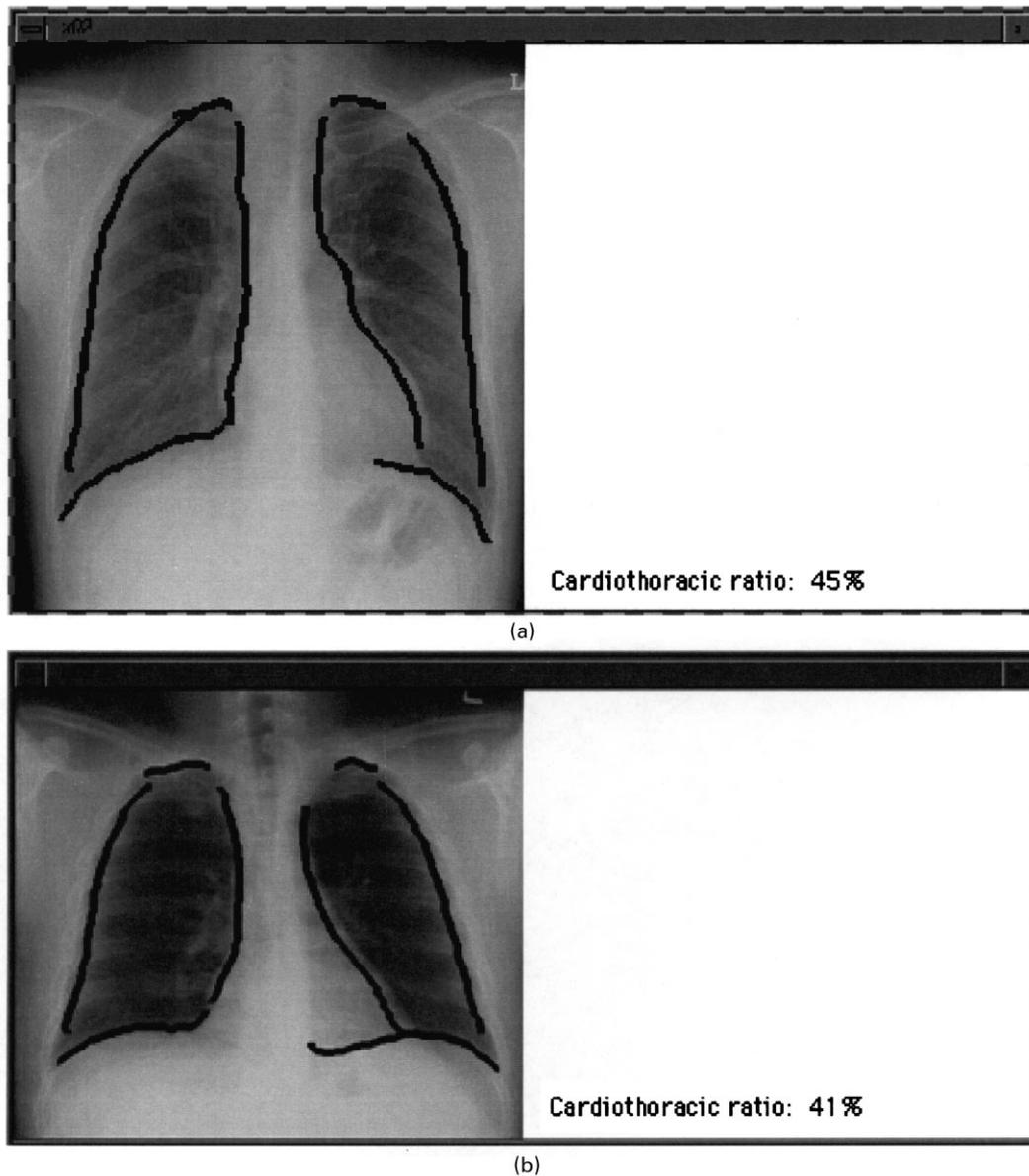


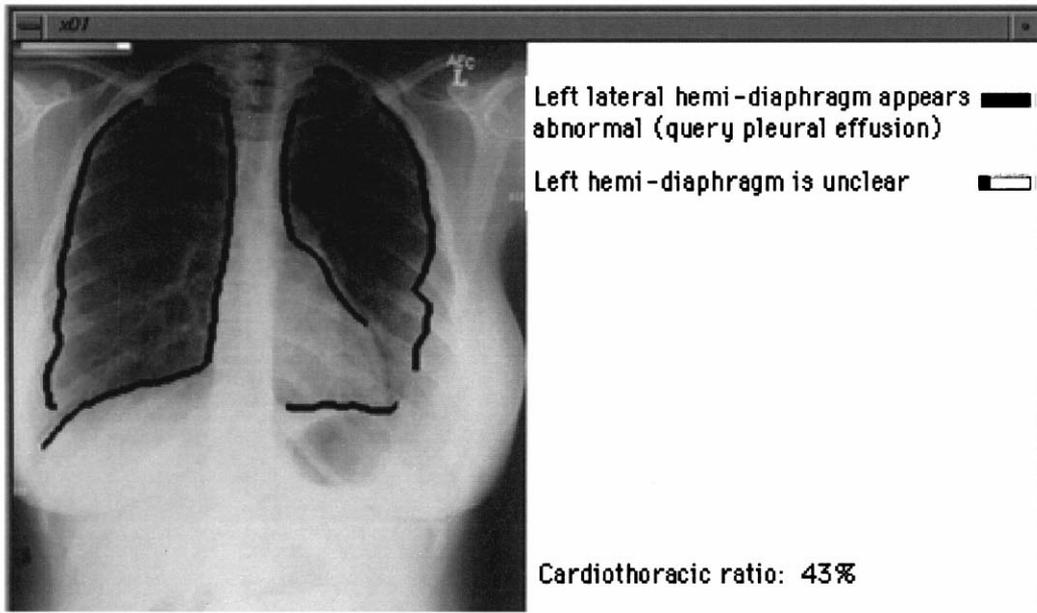
Fig. 12. System output for test images of normal anatomy.

showing no detected abnormalities. Figs. 13 and 14 show labeled edges and automatically generated reports for four of the test images of abnormal anatomy. Most of the reported abnormalities agree with the radiologist's assessment (see Table 4). However, we now examine in detail the causes of a small number of false positive and negative reports from Fig. 14.

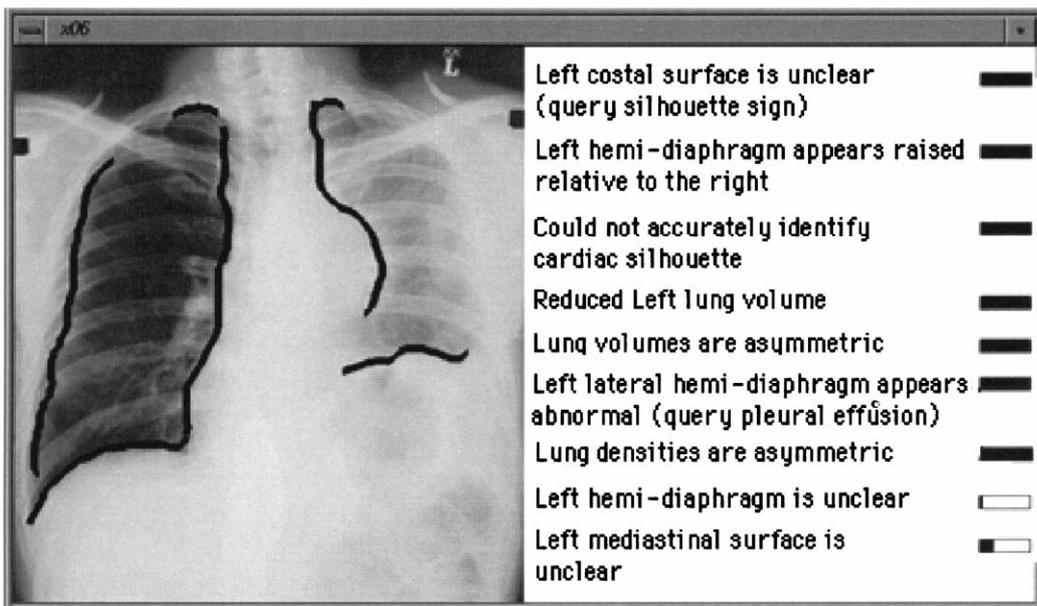
In Fig. 14a, most of the left diaphragmatic silhouette cannot be seen. Therefore the system was unable to identify it and has correctly reported it as "unclear". However, a small amount of the lateral portion of the hemi-diaphragm is visible, and the radiologist was able to conclude that this portion is abnormal and indicative of pleural effusion. Since the model described only the diaphragm as a whole, the system was unable to recognize and report on the small

lateral portion. A more detailed anatomical model may have prevented this false negative. The right hemi-diaphragm has a clearer silhouette (edge) and was correctly identified by the system as exhibiting an abnormality caused by pleural effusion.

In Fig. 14b, the left lung field is totally opaque, and none of its interfaces are visible. However, the system has wrongly identified a rib-edge as the left hemi-diaphragm. This is because the matching strategy is to find the best candidate which meets the model constraints, even if they are satisfied with low confidence. The reason for this approach is to allow the system to recognize highly abnormal (unexpected) anatomy. The mis-identification of the left hemi-diaphragm has led to false positive reports regarding the lung size and raised position of the hemi-diaphragm. The



(a)



(b)

Fig. 13. System output for test images of abnormal anatomy. The bars indicate confidence levels (fully blackened \Rightarrow confidence = 1.0).

system has reported that the labeled edge is both unclear and malpositioned (raised) with respect to its expectations for the left hemi-diaphragm.

5. Discussion

The experimental testing is not intended to be a validation of the system. There are too few subjects, and the test image set was not to a random sample. However, the results illustrate the potential of the system and knowledge-based

methodology. Given the simplicity of the current model, the number of abnormalities that could be identified was encouraging. However, to be clinically useful the system would need to be expanded to include more subtle structures, for example, pulmonary nodules, also much more extensive validation would be required. Nevertheless the current system could be combined with algorithms which specifically detect nodules [5,6] or perform texture analysis [6,26]. Our system could identify the lung boundary and report on whether it was normal and well-defined and thus suitable for further low-level processing.

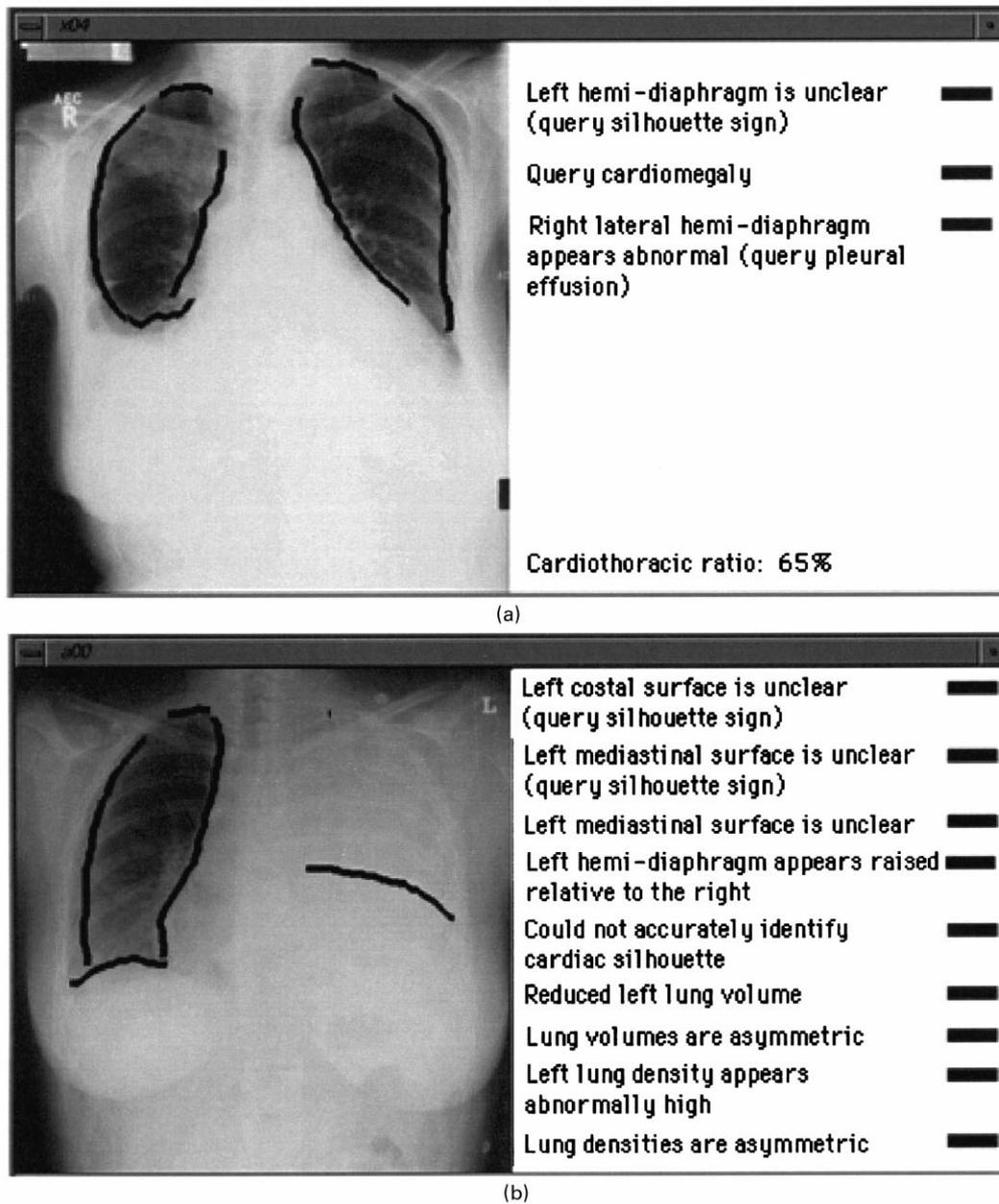


Fig. 14. System output for test images of abnormal anatomy. The bars indicate confidence levels (fully blackened \Rightarrow confidence = 1.0).

The difficulty of the object recognition problem is largely determined by the accuracy of the low-level segmentation routines and the ability of selected features to reliably distinguish the anatomical structures. These factors will influence the amount of knowledge that is required, the number of image candidates, and the level of uncertainty in matching.

The design of the system was motivated by a desire to develop a general-purpose framework for knowledge-based medical image interpretation. The aspects cited as contributions in Section 1 are potentially useful in other applications. The edge detection algorithms and feature-space representation (choice of features used for matching) are specific to projection X-ray images. Elsewhere we have

reported on the application of this methodology to segmenting thoracic CT images [43]. However, application to other problems will be needed before we can conclude that the architecture is generic.

The mapping of the model and image to a common feature space was emphasized because of its importance in giving modularity, and independence of data representations. In this experimental system, the parameters used in the model were chosen to be similar to those in the feature space for simplicity. However, the model could be extended by adding data slots and/or frames. A goal of future work is to develop a more complete and general 3D model [44], but for the chest X-ray application, transformation of the 3D model into the 2D feature-space is a difficult problem.

6. Conclusion

We have developed a knowledge-based approach to lung boundary interpretation in chest X-rays. Within a modular system architecture, an explicit anatomical model is matched to image data by mapping both to a common feature space for comparison. The knowledge-based approach augments low-level segmentation techniques by allowing high-level image interpretation.

In our approach, domain knowledge provides guidance for object recognition. Using the hierarchy implied by relationships in the model, the inferencing and control system automatically schedules the identification of anatomical structures. Both a priori and a posteriori information are combined to constrain segmentation of expected anatomy.

We have developed methods for modeling normal and pathological variations in anatomy. Fuzzy sets provide an intuitive representation and allow symbolic description of image feature values, so that high-level rules can be used to generate reports on suspected abnormalities.

Preliminary experimental results are encouraging but further validation is required.

Acknowledgements

The authors thank our colleagues at the CSIRO, Australia, in particular, Drs. George Kossoff, David Robinson, Habib Talhami, Thanasis Loupas, Mark Berman, Ed Breen, Leanne Bischof and Rosemary Irrgang. Thanks also to Prof. John Hiller of the University of New South Wales and Dr. Michael McNitt-Gray from the UCLA Department of Radiological Sciences for their support and review of the manuscript. Image data was supplied by the Department of Radiology, St. Vincents Hospital, Sydney, Australia and the Department of Radiological Sciences, UCLA School of Medicine.

References

- [1] Katsuragawa S, Doi K, MacMahon H, Nakamori N, Sasaki Y, Fennessy JJ. Quantitative computer-aided analysis of lung texture in chest radiographs. *Radiographics* 1990;10(2):257–269.
- [2] Vargas-Voracek R, Floyd CE. Computer-aided diagnosis of interstitial lung disease: A probabilistic texture-based analysis. *Radiology* 1995;425:197.
- [3] Sanada S, Doi K, MacMahon H. Image feature analysis and computer-aided diagnosis in digital radiography: automated detection of pneumothorax in chest images. *Medical Physics* 1992;19(5):1153–1160.
- [4] Nakamori N, Doi K, MacMahon H, Sasaki Y, Montner S. Effect of heart-size parameters computed from digital chest radiographs on detection of cardiomegaly: potential usefulness for computer-aided diagnosis. *Investigative Radiology* 1991;26:546–550.
- [5] Ballard DH, Sklansky J. A ladder-structured decision tree for recognizing tumors in chest radiographs. *IEEE Transactions on Computers* 1976;C-25(5):503–513.
- [6] Giger ML, Doi K, MacMahon H, Metz CE, Yin FF. Pulmonary nodules: computer-aided detection in digital chest images. *Radiographics* 1990;10(1):41–51.
- [7] Stansfield SA. ANGY: A rule-based expert system for automatic segmentation of coronary vessels from digital subtracted angiograms. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1986;PAMI-8(2):188–199.
- [8] Vernazza GL, Serpico SB, Dellepiane SG. A knowledge-based system for biomedical image processing and recognitions. *IEEE Transactions on Circuits and Systems* 1987;CAS-34(11):1399–1416.
- [9] Fox J, Walker N. Knowledge-based interpretation of medical images. NATO ASI Series, Mathematics and Computer Science in Medical Imaging, Vol. F39, 1988.
- [10] Karssemeijer N, Van Erning L, Eijkman E. Recognition of organs in CT-image sequences: a model guided approach. *Computers and Biomedical Research* 1988;21:434–438.
- [11] Smets C, van de Werf F, Suetens P, Marchal G. An expert system for the labeling and 3D reconstruction of the coronary arteries from two projections. *International Journal of Cardiac Imaging* 1990;5:145–154.
- [12] Delaere D, Smets C, Suetens P, Marchal G. Knowledge-based system for the three-dimensional reconstruction of blood vessels from two angiographic projections. *MBEC North Sea Special Feature*, 1991.
- [13] Dellepiane S, Venturi G, Vernazza G. Model generation and model matching of real images by a fuzzy approach. *Pattern Recognition* 1991;25(2):115–137.
- [14] Garreau M, Coatrieux JL, Collorec R, Chardenon C. A knowledge-based approach for 3-D reconstruction and labelling of vascular networks from biplane angiographic projections. *IEEE Transactions on Medical Imaging* 1991;10(2):122–131.
- [15] Robinson GP, Colchester AC, Griffin LD. Model-based recognition of anatomical objects from medical images. In: Barrett HH, Gmitro AF, editors. *Information processing in medical imaging, Proc. 13th International Conference, IPMI '93*. Springer, 1993.
- [16] Levitt TS, Hedgcock MW, Dye JW, Johnston SE, Shadle VM, Vosky D. Bayesian inference for model-based segmentation of computed radiographs of the hand. *Artificial Intelligence in Medicine* 1993;5(4):365–387.
- [17] Li H, Deklerck R, De Cuyper B, Hermanus A, Nyssen E, Cornelis J. Object recognition in brain CT-scans: knowledge-based fusion of data from multiple feature extractors. *IEEE Transactions on Medical Imaging* 1995;14(2):212–229.
- [18] Benn DK, Pettigrew JC, Sim M, Laine AF, Mancuso AA, De Bose CD, Stambuck HE. A pilot study of automated knowledge-based recognition of facial bones from axial CT slices. In: Lemke HU, Inamura K, Jaffe CC, Vannier MW, editors. *Proc. Computer Assisted Radiology Car '95*. Berlin: Springer, 1995.
- [19] Park W, Hoffman EA, Sonka M. Fuzzy logic approach to extraction of intrathoracic airway trees from three-dimensional CT images. In: Loew MH, Hanson KM, editors. *Proc. SPIE, Medical Imaging 1996: Image Processing, Vol. 2710*, 1996, pp. 210–219.
- [20] Sonka M, Tadikonda SK, Collins SM. Knowledge-based interpretation of MR Brain Images. *IEEE Transactions on Medical Imaging* 1996;15(4):443–452.
- [21] Duryea J, Boone JM. A fully automated algorithm for the segmentation of lung fields on digital chest radiographic images. *Medical Physics* 1995;22(2):183–191.
- [22] Sutton RN, Hall EL. Texture measures for automatic classification of pulmonary disease. *IEEE Transactions on Computers* 1972;21:667–676.
- [23] McNitt-Gray MF, Huang HK, Sayre JW. Feature selection in the pattern classification problem of digital chest radiograph segmentation. *IEEE Transactions on Medical Imaging* 1995;14(3):537–547.
- [24] Toriwaki J, Suenaga Y, Negoro T, Fukumura T. Pattern recognition of chest X-ray images. *Computer Graphics and Image Processing* 1973;2:252–271.
- [25] Chien YP, Fu K. Recognition of X-ray picture patterns. *IEEE Transactions on Systems, Man and Cybernetics* 1974;SMC-4(2):145–156.

- [26] Doi K, Giger ML, et al. Computer-aided diagnosis: development of automated schemes for quantitative analysis of radiographic images. *Seminars in Ultrasound, CT and MRI* 1992;13(2):150–152.
- [27] Armato SG, Giger ML, MacMahon H. Automated lung segmentation in digitized posteroanterior chest radiographs. *Academic Radiology* 1998;5:245–255.
- [28] Hanson AR, Riseman EM. VISIONS—a computer system for interpreting scenes. In: Hanson AR, Riseman EM, editors. *Computer vision systems*, New York: Academic Press, 1978. pp. 303–333.
- [29] Nagao M, Matsuyama T. A structural analysis of complex aerial photographs. New York: Plenum Press, 1980.
- [30] Draper BA, Collins RT, Brolio J, Hanson AR, Riseman EM. Issues in the development of a blackboard-based schema system for image understanding. In: Englemore RS, Morgan AJ, editors. *Blackboard systems*, Reading, MA: Addison-Wesley, 1988. pp. 189–218.
- [31] Kuan D, Shariat H, Dutta K. Constraint-based image understanding system for aerial imagery interpretation. In: *Proc. Annual AI Systems in Government Conference*, Washington DC, 1989:141–147.
- [32] Kulikowski CA. Artificial intelligence methods and systems for medical consultations. *IEEE Transactions on pattern analysis and Machine Intelligence* 1980;PAMI-2(5):464–476.
- [33] Shortliffe EH. *Computer-based medical consultations: MYCIN*. New York: Elsevier, 1976.
- [34] Aikins JS, Kunz JC, Shortliffe EH, Fallat RJ. PUFF: An expert system for interpretation of pulmonary function data. *Computers and Biomedical Research* 1983;16:199–208.
- [35] Nii HP. Blackboard systems: the blackboard model of problem solving and the evolution of blackboard architectures. In: *The AI Magazine*, 1986:38–53.
- [36] Bajcsy R, Lieberman R, Reivich M. A computerized system for the elastic matching of deformed radiographic images to idealized atlas images. *Journal of Computer Assisted Tomography* 1983;7(4):618–625.
- [37] Arata LK, Dhawan AP, Broderick JP, Gaskil-Shipley MF, Levy AV, Volkow ND. Three-dimensional anatomical model-based segmentation of MR brain images through principal axes registration. *IEEE Transactions on Biomedical Engineering* 1995;42(11):1069–1078.
- [38] Kamber M, Shinghal R, Collins DL, Francis GS, Evans AC. Model-based 3-D segmentation of multiple sclerosis lesions in magnetic resonance brain images. *IEEE Transactions on Medical Imaging* 1995;14(3):442–453.
- [39] Minsky M. A framework for representing knowledge. In: Winston PH, editor. *The psychology of computer vision*, New York: McGraw-Hill Book Company, 1975. pp. 211–277.
- [40] Quillian MR. Semantic memory. In: Minsky MM, editor. *Semantic information processing*, Cambridge, MA: MIT Press, 1968.
- [41] Zadeh LA. The concept of a linguistic variable and its application to approximate reasoning-I. *Information and Control* 1975;8:199–249.
- [42] Buckley JJ, Siler W, Tucker D. A fuzzy expert system. *Fuzzy Sets and Systems* 1986;20:1–16.
- [43] Brown MS, McNitt-Gray MF, Mankovich NJ, Goldin JG, Hiller J, Wilson LS, Aberle DR. Method for segmenting chest CT image data using an anatomical model: preliminary results. *IEEE Transactions on Medical Imaging* 1997;16(6):828–839.
- [44] Wilson L, Brown M, Talhami H, Gill R, Sun C, Doust B. Medical image understanding using anatomical models: application to chest X-rays. In: Bizais Y, Barillot C, Di Paola R, editors. *Information Processing in Medical Imaging*. Kluwer, 1995. p. 239–50.

Matthew Brown received his PhD in computer science from the University of New South Wales in 1996. His PhD research was in knowledge-based interpretation of medical images, undertaken in collaboration with the Division of Telecommunications and Industrial Physics of Australia's Commonwealth Scientific and Industrial Research Organisation (CSIRO). In 1998 he became a Visiting Assistant Professor in the Department of Radiological Sciences at the University of California, Los Angeles, where he was continuing research in knowledge-based segmentation and analysis of medical images.

Laurence Wilson received his Physics PhD in 1973 from the University of Sydney. After postdoctoral work at Oxford University, he joined the Ultrasonics Institute, which has now been incorporated in to the Division of Telecommunications and Industrial Physics of Australia's Commonwealth Scientific and Industrial Research Organisation (CSIRO). Initially, he worked in medical ultrasound, particularly tissue characterization and Doppler, where he did research in multidimensional and broad-band methods. More recently, he is leading a group working in computer aided diagnosis, with particular interest in mammography and segmentation using anatomical models.

Bruce Doust received his medical qualifications from the University of Sydney in 1963 and obtained his postgraduate qualifications in Radiology at Sydney Hospital. During the five years at the University of Michigan and the seven years at the Medical College of Wisconsin, he developed particular interests in ultrasound and CT. He returned to Australia in 1981 to take up his current position as Director of Radiology at St. Vincents Hospital and Associate Professor of Medicine at the University of New South Wales.

Robert Gill received his undergraduate education in science and engineering at the University of Sydney and a PhD in electrical engineering from Stanford University. In 1975 he returned to Australia to take up a research position in the Ultrasonics Institute of the Commonwealth Department of Health. In 1989 the Institute was transferred to the CSIRO, Australia's premier scientific research organisation. Dr. Gill is General Manager for Health Services in the CSIRO's Division of Telecommunications & Industrial Physics. For many years his primary field of interest was Doppler ultrasound, with particular emphasis on the measurement of blood flow. More recently he has become involved in the area of computer-aided diagnosis, decision support and knowledge-based image analysis.

Changming Sun received his PhD in the area of Computer Vision at Imperial College of Science, Technology and Medicine, London in 1992. Then he joined CSIRO Mathematical and Information Sciences, Australia, in December 1992 as Research Scientist, both doing research and working on applied projects. His research interests include computer vision, image analysis and photogrammetry. Dr Sun is a member of IEEE, SPIE and The Australian Pattern Recognition Society.