KNOWLEDGE-BASED IMAGE UNDERSTANDING AND CLASSIFICATION SYSTEM FOR MEDICAL IMAGE DATABASES

By

Hui Luo

September, 2001

A dissertation submitted to the Faculty of Graduate School of
State University of New York at Buffalo
In partial fulfillment of the requirement for the degree of
Doctor of Philosophy

Department of Computer Science and Engineering

In memory of my mother

Nanxiu Huang

Abstract

The understanding of medical images involves the identification of anatomical structures in images, the establishment of the relationships among the structures and the matching them with pre-defined anatomical models. The closer this match, the better the image is understood. Hence, the development and use of anatomical knowledge models are critical to the effectiveness of the system. In this thesis, we will present a new knowledge model for medical image content analysis based on the object-oriented paradigm. The new model abstracts common properties from different types of medical images by using three attributes: description, component, and semantic graph, and specifies its actions to schedule the detection procedure, properly deform the model shapes to match the corresponding anatomies in images, select the best match candidates, and verify the candidates' spatial relations with the semantic graph defined in the model. The advantage of the new model includes: 1) It presents a robust representation ability to represent the object variability, 2) It provides a tighter bond between features and methods, and 3) It supports the complex image analysis and works on both normal and abnormal images.

Based on this model, a knowledge-based medical image classification system is developed. The system assigns each input image with one or more most similar models by performing a two-stage processing. The first stage, coarse shape classification, focuses on the global shape features of objects in the processed images. The second stage performs a detail matching between extracted image features and model specified properties. The output match results imply a set of models which are most possible to be existed in the processed image. To further confirm them, an evaluation strategy is used to reject those unreasonable results and inference the high confidence match models.

To improve the detection accuracy, we also proposed a new region model which combines both region and edge information into image segmentation, and reformulate the traditional snake model and level set model with it. As expected, the reformulated models demonstrate robust ability to deal with the ambiguity in images and overcome the problem associated with the initialization.

The performance of the system has been tested on different types of digital radiographs, including chest, pelvis, ankle, elbow and etc. The experimental results suggest that the system prototype is applicable and robust, and the accuracy of the system is nearly 70% in our image databases.

Key words: Medical images databases, Knowledge-based, image understanding, image classification, feature extraction, snake model, object-oriented representation

Acknowledgements

I would like to thank my thesis advisors, Dr. Acharya and Dr. Gaborski, for all of their guidance, encouragement and support in my research and overall graduate experience, without which this dissertation would not be possible.

Four years ago, Dr. Acharya and Dr. Gaborski introduced me to the field of medical image analysis, and I immediately found my interests in this area. Over the years, they have given me the help to formulate my research problems, the freedom to work on the problems, the direction to help me stay aware of the big picture and evaluate benefits and pitfalls of various solutions I have considered.

I also thank Professor Aidong Zhang, Professor Venu Govindaraju, and Professor Rohini Sridhar for serving in my committee, and giving me many valuable advice and direction.

Many thanks to the thesis outside reader Professor Zhongfei Zhang for his suggestions and comments, which improve the final draft of the thesis.

My special thanks go to Health Imaging Lab of Kodak Company and Siemens Corporate Research for providing financial supports at various stage of my thesis.

Over the passed four years, I have the opportunity to work at the Networked Multimedia and Visualization Laboratory, a wonderful place. I want to thank all former and current members at the lab for helping me during these years, giving me friendship, and making my stay at Buffalo pleasant and memorable.

Most importantly, I would like to express my gratitude to my family. This thesis is dedicated to my mother for her love and encouragement, with which I always have courage to face difficulties in my life. I also thank my father, who showed me the path of academics where I walked this far, and my lovely sister for her love and encouragement. Special thanks go to my husband for discussing me with my ideas, sharing my difficulties, and always understanding and supporting me.

Contents

3. Robust Snake Model 50

List of Figures

List of Tables

Chapter 1:

Introduction

1.1 Significance

Medical images, at the heart of diagnosis, therapy treatment, and surgical planning, have undergone tremendous changes in the past two decades. Today they not only have various image modalities (e.g. X-ray Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Ultrasound (US), Position Emission Tomography (PET), Single Photon Emission Computed Tomography (SPECT), Digital Subtraction Angiography (DSA), and Computed Radiograph (CR), etc.) but also can be stored into digital format for further processing. The amount of digital medical image data captured per year in U.S. alone is estimated on the order of petabytes –10^{15} bytes –and increases rapidly every year [Wong 97].

The related study of medical images is shifting from generation and acquisition of images to post-processing and management of medical images, using multimedia clinical information and computer-assisted diagnosis. Among most current research topics in this field, archiving and managing huge volumes of medical images have become more and more important, because it is core in health care information technique which is one of the fast growing segments of the health care market. In particular, the Internet is enabling better communication and collaboration between patients and doctors, which will result in the significant improvement of healthcare as well as the reduction of costs.

The direct benefits of medical information management system includes:

- Post-processing

  As a part of medical information management system, medical image classification plays an important role in automatically, intelligently and efficiently post processing images before presenting them to the human user. For example, through retrieving and comparing the similar images from medical image databases, the system enable to automatically set the processed image display parameters, determine the post-processing filters, and set corresponding image acquisition parameters.

- Aiding in diagnosis

  With the help of medical information management system, we can easily retrieve related patient studies from the databases for comparative evaluation, validation of diagnosis, and judgment of effectiveness of therapy.

- Teaching and training tools

  Continuous education and self-assessment are very important in medicine domain. Online availability of medical databases opens new avenues for aiding diagnosis, training and teaching. Medical information management forms an important part of implementing flexible teaching tools

- Telemedicine

Medical information management plays a key role in building easy-to-use telemedicine system, by which physicians in small clinics or remote sites

can benefit from accessing databases in expert hospitals.

1.2 Research Scope and Goal

The goal of this research is to provide a systematic framework to automatically analyze and classify medical images. Due to the limitation of image sources, the medical image database used in this research only contains radiographs. However, the designed framework is not constrained to radiographs, it enables to be extended to other image modalities.

To accomplish the goal, the knowledge of the relevant local anatomy is commonly introduced into image analysis. In the field of artificial intelligence and knowledge-based computer vision, the choice of knowledge representation is very important to ensure good solutions, but it is still a major open issue. Thus one specific aim in our study is to propose an efficient and effective way to represent anatomical knowledge and then combine it with image process routines to classify radiographs.

Due to the poor quality of radiographs and the large variability of anatomy structures of the human body, Using a single low-level image segmentation technique is hard to generate the satisfied results. So it is quite reasonable and necessary to take advantage of all useful information obtained from different techniques. However, different segmentation techniques will generate different results. In such a situation, the priori knowledge will be useful to resolve the differences between various techniques. So, as another specific aim, we design a new image segmentation method which benefits from the rich information provided by both image data and the priori knowledge.

1.3 Contribution

We view the contribution of this research in the following areas:

1) Computer Vision and Image Processing:

- *Proposing a new object-oriented knowledge model for image analysis:* the proposed knowledge model provides a framework for abstracting the complex image properties at different levels and coping them with the corresponding vision operations in complex image analysis based on the object-oriented idea.
- *Reformulating snake models:* Aim to overcome the problems suffered by the traditional snake model, the new model demonstrates robust ability in dealing with the ambiguity in images by introducing a region model which combines both edge and region information into segmentation.

2) Medical Imaging:

- *Designing the architecture of medical image classification system:* To simulate human perception behavior, the proposed system is designed as a two-stage hierarchical structure. It first focuses on the coarse shape outline of the main object in an image, then goes into the detail information of contents. The whole procedure combines both image information and anatomical knowledge by using an object-oriented knowledge model.

1.4 Organization of the thesis

We organize this thesis as follows:

Chapter 2 proposes a knowledge model for medical image content analysis based on the object-oriented paradigm. The new model abstracts common properties from different types of medical images by using three attributes: description, component, and semantic graph, and specifies its actions to schedule the detection procedure, properly deform the model shapes to match the corresponding anatomies in images, select the best match candidates, and verify the candidates' spatial relations with the semantic graph defined in the model.

Chapter 3 investigates the properties of the snake model in detail and discusses initialization and implementation with respect to the energy minimization framework of the snake model. With the motivation of developing new improved techniques, a new region model is promoted to combine region and edge information in the segmentation, which directly overcome the problems associated with the initialization. Based on the proposed region model, we reformulate the traditional snake model and level set models.

Chapter 4 presents the architecture of knowledge-based medical image classification system, which employs both global and local features to classify the input images by performing a two-stage processing. The first stage focuses on the global shape features of objects in processed images. The second stage employs local features to identify salient component objects in images and incorporates their spatial relation to discard those spurious match candidates.

Finally, Chapter 5 summarizes the work and suggests future research direction.

Chapter 2:

Design of the Object-Oriented Knowledge Model for Image Understanding

2.1 Introduction

The goal of image understanding system often involves the identification of objects in images and the establishment of the relationships among the objects. This transformation of signals (the images) to symbols (the interpretation) in the visual domain is one of the most important and most difficult task in artificial intelligence. There is no working system yet which comes close to the capabilities of the human visual system. Some reasons are:

- The vision problem is ill-defined, and its terminology has been under a constant change.
- The way with which human capture image contents, group image features into meaningful objects and attach semantic descriptions to images has not been fully

understood to automate the image analysis procedure.

- Problems arisen in segmenting an image into regions corresponding to individual objects, extracting features from the image that capture the perceptual and semantic meanings and matching the image with the proposed model based on the extracted features make the analysis problem more challenging.
- The needed computational power for real-time digital image analysis exceeds even the best workstation.
- Specialized problem solving methods can hardly be generalized.

2.1.1 Our Motivation

The algorithmic approach of processing and analyzing digital images has resulted in powerful means to interpret specified kinds of images, but failed in providing general image understanding methods that can work on all kinds of images. This is mostly due to the fact that every image can be interpreted in many ways, as long as we don't know what we expect to see in it. Thus as a main goal of this research, we need to build a model about the expected contents of images in order to be able to "understand" them.

Another important goal is to provide a theoretical framework in the model which appropriately integrates multiple special-purpose image algorithms and techniques with the knowledge base.

This remainder of this chapter is organized as follows: Section 2.2 reviews some knowledge representation schemes used in image understanding. Section 2.3 introduces some ideas to deal with the complexity in image understanding. Based on these ideas, the basic concepts of the proposed object-oriented knowledge model are defined in Section 2.4. Section 2.5 details how to implement the knowledge model into images analysis and some case studies will be given. Finally, a conclusion is drawn in section 2.6.

2.2 Overview of Knowledge Representation in Image Understanding

The analysis of images involves the matching of features extracted from the image with pre-defined models. The closer this match, the better the image scene is understood. Hence, the development and use of these models are critical to the effectiveness of the system. A good production of a high-level model requires the representation of knowledge about the object characteristics, relationships and when and how to use the information stored within the model to process the image data. In this section, we will review several typical knowledge representation schemes which have been developed in AI and been introduced in image understanding. The motivation of the discussion is to determine the emphasis for developing the new knowledge model.

2.2.1 Production Rules System

In the production rules system, knowledge is encapsulated in a set of IF-THEN rules or condition-action pairs which interact with image data. The previous work for rule-based image analysis systems mainly focus on low-level image segmentation, such as using a set of production rules to generate image processing flow sequences [Nazif 84, Brook 81], evaluating the preliminary segmentation results [Lin 88, Mckeown 87] etc. Because of the complexity of images, a large number of rules are needed in a real world task domain, also the additional control rule are indispensable and have usually been added by structuring the rule base into classes or phases. Unfortunately, rules are not natural representations for many declarative facts required in the knowledge base. The rule-based systems by themselves do not guarantee addition of knowledge and consistency reasoning [Binford 82]. In addition, these approaches defeat modularity and tend to make program modification difficult [Draper 89].

2.2.2 Blackboard Systems

The framework of a blackboard system includes: (1) a blackboard, (2) knowledge sources and (3) a control scheme. The blackboard represents a common database through which all knowledge sources communicate. The knowledge sources contain the partial knowledge required to solve the pieces of the problem and respond opportunistically to the data posed on the blackboard. The control scheme is charged with determining which of the knowledge sources is best suited to contribute next to the emerging solution of the problem at hand. This blackboard architecture provides a means for implementing opportunistic reasoning which is a useful problem-solving paradigm to be employed when the combination of various sources of knowledge is necessary to solve a highly complex problem. Some recent blackboard type systems are described in [Levine 81, Matsuyana 89]. The common feature of them is the working memory which acted as repository for both the input data and the deductions made by the different image processing modules. Draper et al [Draper 89] pointed out two problems with blackboard architectures in respect to image processing. One comes from their control structure through which the blackboard decides which image analysis processes will run and consume resources. With large number of knowledge sources involved in image processing, the control efficiency becomes problematic. The other difficulty lies in the temporal continuity of image knowledge sources, blackboards assume that various knowledge sources will wake up when required for use and then vanish. In image analysis, knowledge sources tend to be associated with objects or regions of images, which persist as long as the image remains under analysis.

2.2.3. Frame-based Semantic Networks

A *semantic network* is a labeled, directed graph. The network nodes are used to represent objects in images, and the arcs connecting the nodes represent various relationships between the objects. Both the nodes and the arcs are labeled by attributes corresponding to properties of objects and relationships, respectively. This type of semantic network can describe complex relations of objects explicitly and succinctly, and also reduce the search time for particular objects with the well-organized structure. However, the major problem of semantic networks is the lack of theory of semantics. No standard interpretation exists for knowledge expressed within a network, and invalid inference can be drawn from the knowledge contained within the network [Gonzalez 93].

Frames provide the framework for representing knowledge at a more systematic level. As data structures, frames offer the benefits of record structures, slot-access procedures, value inheritance and object hierarchy. Individual frames correspond to the nodes in the semantic network and the relationships between the frames are the arcs in the semantic network. The slots in the frames form a description of the object and also contain production rules for instantiating other object frames when appropriate. In this formalism, frames are thus control mechanisms as well as data representations. The control paradigm of a frame interpreter is essentially two phases: a structure matching phase and a forward-chaining phase [Gonzalez 93]. In the structure matching, a candidate frame instance is chosen and the data is searched in an attempt to fill the empty slots in that frame instance. The forward-chaining phase is initialized by attached demon when a slot is filled. This representation hides the procedural complexity and makes frames clean. Numerous systems using this frame-based semantic network were reported to aid in the analysis of images [Minsky 75]. Hanson and Riseman [Hanson 82] used frames as hypothesis generation mechanisms to predict the presence of objects. Ikeuchi and kanade [Ikeuchi 88] used frames to represent aspect of three-dimensional objects. Walker et. al. [Walker 90] used the frame-based representation's ability to represent partial specified objects to adjust the level of processing to available data. However, Draper et. al. [Draper 89] pointed out that using networks of frames for both control and representation tended to hide the procedural behavior of a system and destroy its temporal coordination. Such a formalism offered " a snapshot of the network of data relations, but ordering, branching and looping behaviors are extremely difficult to specify."

2.2.4 Object-Oriented Representation

An *object* is defined as a collection of information (i.e. the data) representing an entity within the real world and a description of how this information is manipulated (i.e. the associated procedures). Several principles of the object-oriented approach make it well suited for representing knowledge, like abstraction, encapsulation or information hiding, inheritance and polymorphism. Although the object has much in common with frames such as they both support semantic networks, inheritance, abstraction and the concept of procedural attachments, the object-oriented representation is superior to frames in that object methods (procedures) are more universal since they provide any type of general computation needed and support encapsulation and polymorphism. The essential ideas of object is to describe data not by its internal structure, but by its externally observable behavior. This separates the issue of how the data is implemented from what it achieves. While frame demons serve only to compute values for the various slots or to maintain the integrity of the knowledge base whenever an action by one frame affects another. Hence, the object-oriented representation creates a much tighter bond between the data and action than frames [Gonzalez 93].

Although the object-oriented knowledge representation is considered as a natural and effective approach, the use of object-oriented within complex image analysis has not undergone a rapid growth as it happened in other fields. We argue that one of major problem comes form the difficulty of conceiving a comprehensive framework for coping with the different abstraction levels and the vision task operations. With the goal to overcome such a drawback, we present a new knowledge model for processing and analyzing of medical images based on the object-oriented knowledge representation. The model aims to provide an efficient and effective framework to represent the knowledge related with processed images and combine it with image routines for better image understanding.

2.3 Strategies for Dealing with Complexity

Image understanding has to deal with the ***Symbol-to-Signal Gap*** and extract a symbolic description from an image. According to the characteristics of different processes, the image understanding system can be classified into the following four representation levels, as shown in fig 2-1

1. Preprocessing: Improve and normalize the raw image data
2. Segmentation: Extract the relevant image features (regions, edges, features) by top-down or bottom-up analysis
3. Object recognition: Employ the priori knowledge about the expected objects to recognize objects, and extract a symbolic description to the objects
4. Image understanding: Combine the knowledge base and the processed results to interpret the image scene.

As can be seen, the knowledge sources have an influence on each of these levels. In general, the knowledge based image analysis systems have to face three fundamental issues: what knowledge relates to a problem, how that knowledge should be represented and how and when that knowledge should be used. Due to these issues, the design of knowledge-based image understanding systems is more complex than that of traditional image analysis systems. In this section, we examine some of the techniques for dealing with complexity and analyze how to embed them into our object-oriented image analysis knowledge model design.

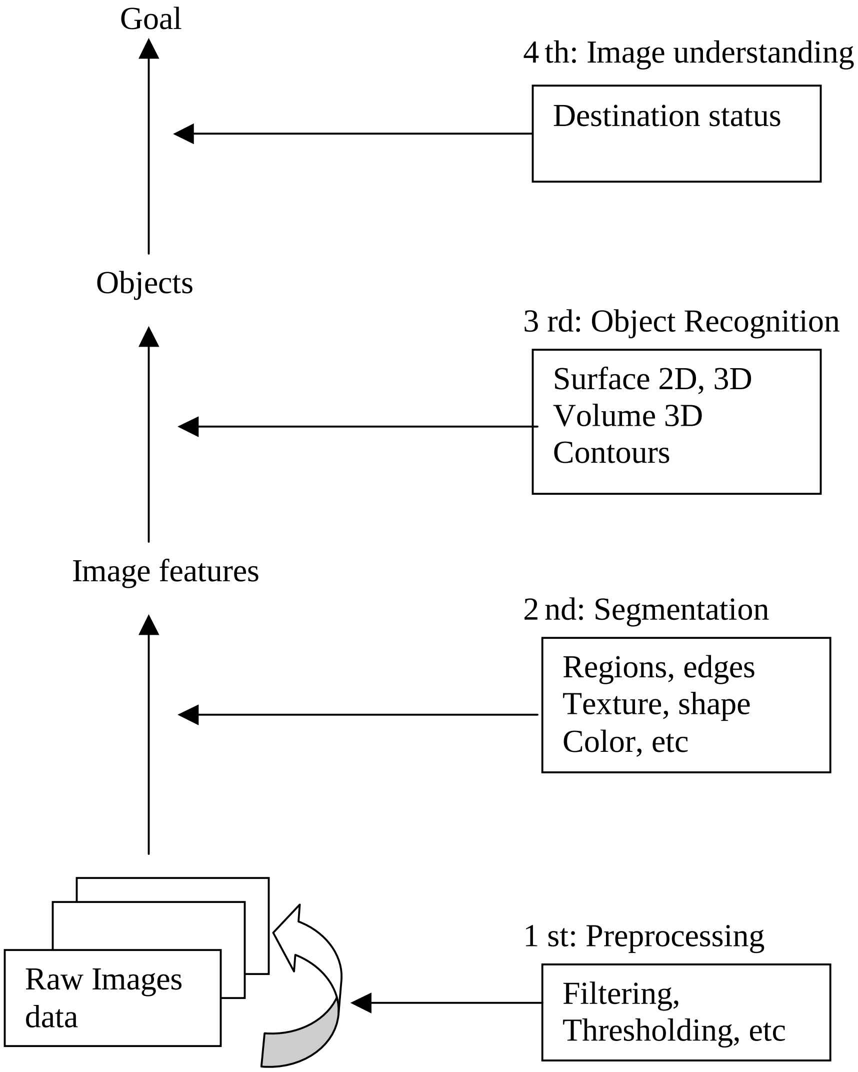


Fig 2-1: General levels in an image analysis system

2.3.1 Abstraction

The most common and effective technique for dealing with complexity is abstraction, which ignoring aspects of some entities that are not relevant to the current problem and concentrating more on those relevant aspects. General speaking, the abstraction encourages both procedural and data abstraction. In procedural abstraction, an operation can be viewed as a single entity even though it is actually achieved through some sequence of low-level operations. Data abstraction is similar. An object is viewed as a single object rather than as a composite of all of its individual parts. This suppression of the lower-level details allows us to reason about the operation or object at hand more efficiently. However, the major problems in abstraction are how to determine levels of abstraction and which aspects are significant at each level.

Based on the properties of representation levels in the image analysis system, we classify the knowledge base into two levels of abstraction: the detail level and the schematic level.

The detail level directly deals with image data and focuses more on low-level image processing procedures, such as image preprocessing, segmentation and object recognition. Thus, it includes concepts for modeling the expected object structure and behavior, such as how to specify the object properties, associations for the object relationship, and how to design the corresponding control schemes and efficient match strategies to detect the objects in images.

The schematic level only concentrates on high-level analysis and tries to extract a symbolic description to the processed image. So, it uses the results from the detail level as inputs and provides concepts for dealing with higher-level interpretation, as well as concepts for describing high-level behavior to accomplish the goal.

2.3.2 Hierarchical Representation

Since there is a large representation gap between the initial image data and the high-level symbolic descriptions that constitute an interpretation, hierarchical representation tries to establish multiple correspondence levels between image features and the knowledge base based on the intermediate representation of an interpretation.

For a number of reasons, the image interpretation cannot be accomplished by one or few procedures across a wide range of image contents, but rather performed via a sequence of partial interpretation of image contents. This is not only because of the complexity of image objects, but also because of analysis goals and the priori knowledge used in the image analysis. A study of human vision system [Aggarwal and Cai 99] shows that human vision is frequently a goal-directed process, we recognize a complex object in scenes by anticipating that all component of the object are present, and also their relationships are satisfied with what are defined in the priori knowledge. Viewing in this manner, we simplify a complex object representation by using a set of salient component objects. This allows us to abstract out significant features of component objects which are easy to be extracted and the particulars of their relationships, and also encode knowledge of objects and detection vision algorithms as a combination of data structures and interpretation procedures.

The object-oriented approach supports such hierarchical representation. As the basic elements in building the object hierarchy, the selected primitive features are assigned to the lowest layer of the hierarchy which corresponds to the segmentation level in the image analysis system. The objects to be recognized or the images to be understood are defined in the highest layer which corresponds the understanding level in the analysis system. Between these two layers are middle layers, where intermediate objects are created. Intermediate objects refer to components of the object or sub-regions in the image, which are usually generated in the recognition level. This hierarchy also offers a number of hierarchical abstractions, which means the combination of several single detailed objects into an abstract one, as well as the scaling down the number of detail objects. The image interpretation is performed based on extracted primitives and creates all intermediate objects progressively according to the hierarchical object model, layer by layer, until the interested objects are formed or the expected image scene is verified. Hence, the hierarchical representation facilitates the image interpretation of higher complexity.

2.3.3 Control Issue

Image content understanding can be achieved only as a result of cooperation of complex information processing methods and appropriate control of these methods. The task of the control strategy is to find the best sequence of vision procedures to achieve the goal, where “ best” may be defined in terms of accuracy, cost or more often a combination of both. As image information is stored at different representation levels during processing, there is one crucial question related to processing control: should the processing be controlled by image data information(bottom-up control) or by higher-level knowledge(top-down control). In the top-down control, a set of assumptions is constructed from the knowledge base. During image representation, the satisfaction of image features with the pre-defined assumptions is tested at each processing level in a top-down direction, down to the original image data. In such situation, the image understanding is model verification. Performing in a totally different direction, the bottom-up processes carry out on raw image data, look for as much structures as possible without invoking object-specific knowledge, then match the resulting structures with available object knowledge to construct an intermediate representation and finally derive the interpretation from intermediate representation. However, efficiency considerations plead for the early introduction of top-down processes in image understanding. Since first in many situations, large areas of the image simply do not have to be looked at. Second, pure bottom-up processing leads to a combination explosion. Third, contextual knowledge about how objects related to each other provides semantic clues about what and where other objects to look for. Thus an appropriate combination of bottom-up and top-down operations can yield a more flexible and powerful vision control strategy for complex image understanding.

Corresponding to the hierarchical representation, the control strategies in the lower layers aim to detect object primitive features like shape, size, color and texture etc. So the most important issue of the control strategies is how to appropriately cooperate the top-down or bottom-up operations to extract low-level features. In these layers, both contextual knowledge, like how objects are related to each other, and image information are used. If strong contextual information and weak visual data are available in one situation, the top-down process can be used to simplify the problem. Otherwise, it appears that the bottom-up processing would have more accurate results. For the high layer such as the object or image layer, its control strategies concern more about how to divide a complex structure into a set of component structures which have easy detected features, and how to arrange the optimal search path to minimize processing time and maximize the reliability of detection.

2.3.4. Handling Uncertainty

There is a significant amount of inherent ambiguity in image understanding. In order to arrive at an unambiguous final result, the knowledge base must include a sufficiently rich set of constraints and flexible mechanisms for manipulating uncertain hypotheses until there is a convergence of evidence.

The matching processes responsible for generating correspondences between image objects and model objects will inevitably make errors and may generate multiple competing hypotheses to describe the expected model object. These result in uncertainty at all levels of representation. Generally speaking, there are two kinds of ambiguity: imprecision and uncertainty. The imprecision comes from the fuzzy aspect of image objects, since it is difficult to locate an image object perfectly accurate. The uncertainty comes from doubts about the object position, for example, it “may be” here or it “may be” there. To deal with uncertainty, the following three issues must be taken into account:

1) How to represent uncertain data
2) How to combine two or more pieces of uncertain data
3) How to draw inference using uncertain data

These issues involved are often very subtle, but they can have significant impact on the operation of any conclusions drawn by a system.

In this research, fuzzy logic is introduced to handle uncertainty, The theory of fuzzy sets proposed by Zadeh [Zadeh 65] allows an efficient manipulation of ambiguities by using linguistic variables to describe vague concepts like existence of an object, and fuzzy reasoning is well suited to express an priori knowledge as a set of fuzzy rules which determine the future situation on the basis of present observations. Our method transfers a real world attribute value into linguistic variables and uses fuzzy membership functions characterize the possible degree of membership that the attribute value belongs to the linguistic variables. The advantage of this fuzzification is that it provides a reasonable way to represent the imprecision of object features. Then fuzzy reasoning rules are applied to recognize the object by integrating the priori knowledge with their fuzzified attributes. Finally, a confident level is assigned to each object candidate based on how it satisfies the knowledge base and subjects to any relevant constraints or relationships. Only the high confident candidates are selected as the final solution.

2.4 Basic Concepts of the Object-Oriented Knowledge Model

The performance of an image understanding system depends heavily on the design of the knowledge model structure. Thus to design a model which provides enough capability for the general application, the model has to address and solve the following issues [Crevier 97]:

- Adapting to variations among the knowledge model and instances
- Planning processing operations, such as breaking down the main goal into sub-goals and further dividing the sub-goals into processing steps
- Controlling the processing flow during the analysis, for instance bottom-up or top-down processing, distributed or centralized control
- Representing and using knowledge into image processing
- Implementing evaluation mechanisms for image processing quality
- Recognizing objects by matching them with models and resolving conflicts between contradictory hypothesis

As mentioned before, the object-oriented knowledge representation provides a tight relationship between the data and the actions, which actually corresponds more closely to our view of object within the real world and provides a much firmer basis for image processing program development. Also its pleasing properties, such as abstraction, encapsulation or information hiding, and hierarchy can simplify the view of the problem, optimize the processing procedures, and minimize the processing time. Thus the object-oriented class structure is especially suitable for the image analysis knowledge model.

In this section, we will address how to design this object-oriented image analysis knowledge model to satisfy the above concerns. The whole definition of the model is divided into two parts: model attributes and model methods. We start with the conceptual level notion of model class which introduces three attribute parts: description, composition and semantic graph. Then we explicate the model methods which serve as realization mechanisms to link image data with the knowledge model.

2.4.1 Model Attributes

An important requirement in designing a generic knowledge model is the capability to describe variations between a generic model and individual instances. We view this adaptation at two levels: one level is to adapt the generic model to a specific category model, the other level is to adapt a specific category model to an individual instance. To accomplish these, the designed generic model class abstracts common features from images into three attribute parts: the description part, the composition part and the semantic part, as shown in Fig 2-2.

The Description Attribute

The description attributes represent the difference between a model and its instances by using a set of features, such as size, shape, color, texture, position and orientation. For each attribute, the upper and lower "deformability" limits are defined to specify the possible range of values under natural deformation.

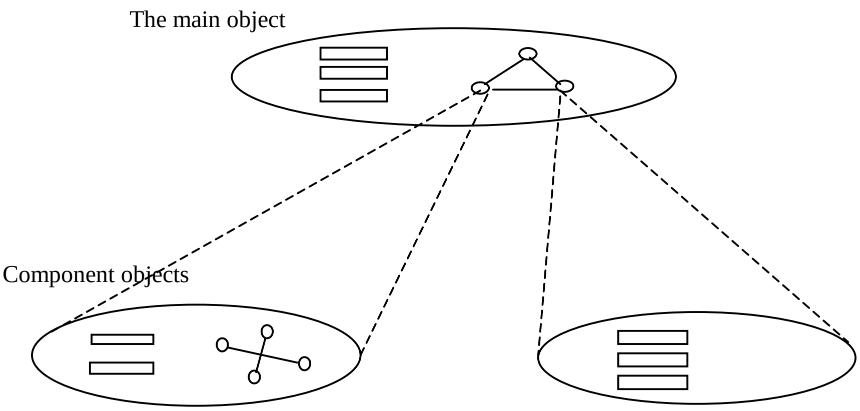


Fig 2.2: Representation of an object-oriented knowledge model

▭ Description attributes
○ Component objects
— Semantic relations

The Composition Attribute

The hierarchical building of an object model is the crucial point of the image understanding, since it not only simplifies the description of structure objects, but also facilitates the image interpretation of higher complexity. To embody this property into the model design, a composition attribute is introduced into the model object to specify all model component objects at its next lower layer. The main object which corresponds to the model object or the processed image scene at the top layer gives a global view of the object or the image. It contains not only the global description attributes such as the object size or orientation, but also a composite attribute to specify a list of all component objects existed in the main object. In such structure abstraction, the main object can be viewed as a single object. For those component objects at the lower layers of the hierarchy, the composition attribute may or may not exist depending on whether the component object is enable to be further decomposed.

The Semantic Graph Attribute

The spatial relationships among component objects outline the structure of an object or an image. This information can be easily encoded in our object-oriented knowledge model by using the semantic graph attribute. In general, there are two properties included in this part. One is the connection between the same-level component objects. The other one is the adjacency referring to the contacts between component objects, such as right of, left of and so on.

These three attribute parts together present a robust representation ability. Since they not only fit to represent the object variability but also are suited to express three common relations: the spatial adjacency relation, the hierarchy relation and the inheritance relation.

2.4.2 Model Methods

This section details the mechanisms by which images can be connected with their corresponding knowledge models. In short, model methods provide strategies to recognize the model objects in images. The model objects in this study imply the main object and all its component objects. Here, we concentrate on three main methods in the model: control strategy, match mechanism and inference engine.

Control Strategy

The overall task of the control strategy is to select computer vision procedures that will provide the most accurate results at the lowest cost. As discussed in section 2.3.3, the control strategies at the high layer can be viewed as a heuristic search problem and aim to search for the optimal sequence of detection procedures. The control strategies in the lower layer are to seek an efficient and effective combination of both bottom-up and top-down process methods to facilitate the feature detection.

Match Mechanism

Matching is another basic model method which attempts to recognize objects by matching image data with known characteristics of an object. Some objects are described and matched by their shapes, while others are matched by features like color, texture, or context. Image understanding system should be capable to select the appropriate features for the current target object and allow to adapt to variations among instances of the object. In general, match processing is time-consuming even in the simplest cases, but the process can be made faster if a good operation sequence is found. An efficient strategy is to look for the best partial matches first, followed by verifying a graph construction of these partial matches. A good example is the recognition of a rectangle in a noise-corrupted image. After segmentation, the broken lines are obtained as the primitives. The searching for full lines as intermediate objects before building the rectangle seems to be a better and more sensible policy than directly building the rectangle from the primitives. Our hierarchical model structure already provides such a good mechanism for this idea.

The match schemes in the component objects implement matching directly with their related image sub-regions. Once the matching of all component objects is finished, the match scheme in the main object combines all match candidates to construct a set of combination graphs of image objects, and then verifies the graphs with the semantic graph attribute in the knowledge model to infer the best match results.

Based on the features selected, the matching criteria may be variant. For some simple features, such as color, size and texture, the match criteria usually compute the distance between instance features and the predefined properties of the corresponded object. The smaller the distance, the better the match. However, for other features like shape and spatial relation, it is hard to define the distance measurement to evaluate their similarity. So the following matching criteria are taken into account.

Matching of the Component Object Shape

The shape matching would be extremely easy if an exact copy of the interested pattern could be expected in the processed image; however, in real images, some parts of the pattern are usually corrupted by noise, geometric distortion, occlusion, and etc. Many facts like size, orientation and distortion will also influence the accuracy of the correlation value. Therefore, it is impossible to look for an exact match, a search for location with optimal match is more appropriate.

The shape matching criteria can be defined in many ways: in particular, correlation between a pattern and the searched image data is a general matching criterion. In this study, we assure that a shape pattern is found in an image only if the correlation between the shape model and the edge map of the image exceed a pre-defined threshold. However, to perform such matching in real world images is not easy as the reasons mentioned before. To overcome such problem and improve the match quality, we properly deform the model shape according to the image features, such as size, orientation, and also relax the model shape constraint by convoluting it with a matrix M. So that the edge of the model is no longer one pixel but at least seven pixels wide, which will greatly increase the matching possibility.

$$M = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

The following formulations (2.1-2.3) presents a matching value between *f* (the edge map of images) and *h* (a component object shape after convolution) located at a position *(u, v)* in the image, and *V* is the set of elements in the component object shape matrix.

$$C_1(u,v) = \frac{1}{\left(\max_{(i,j)\in V} \left| f(i+u, j+v) - h(i,j) \right|\right) + 1} \qquad (2.1)$$

$$C_2(u,v) = \frac{1}{\left(\sum_{(i,j)\in V} \left[f(i+u, j+v) - h(i,j) \right]^2\right) + 1} \qquad (2.2)$$

$$C_3(u,v) = \frac{1}{\left(\sum_{(i,j)\in V} \left| f(i+u, j+v) - h(i,j) \right|\right) + 1} \qquad (2.3)$$

To speed up the process and reduce the processing complexity, we derive a threshold from numerous experiments. Only the positions with the correlation value exceeding the threshold are kept, along with their corresponding correlation values. Then, we construct a label image with the same size of the processed image, in which only the kept positions are labeled by their corresponding correlation values and the rest of the positions are marked as zeros. Because of the spatial continuity of matching results, it is easy to observe that there are several clusters existed in the label image. Among these clusters, we are only interested in the position with the maximal correlation value in each cluster. One simple way to achieve it is to perform the Non-maximal suppression algorithm as described below:

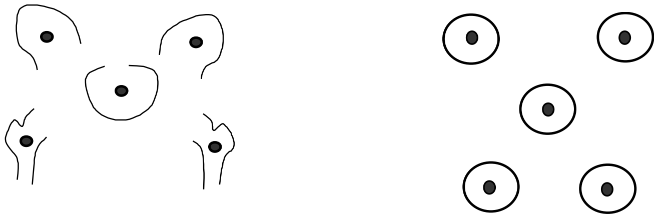
Non-Maximal Suppression algorithm

1. *For each pixel with non-zero correlation value in the label image, inspect its 8-connectivity.*
2. *If the correlation value in one of its 8-connectivity exceeds that of the pixel under inspection, mark the pixel for deletion*
3. *When all pixels have been inspected, rescan the label image and erase all pixels marked for deletion to zero.*

The hypothesis used here is that: if a processed image contains the expected shape pattern, there will exist at least one correlation value which is greater than the pre-defined threshold. Otherwise, the sub-image doesn't contain the expected shape pattern.

Matching of the Semantic Graph

Once a component object is found in the image, we may have one or more match positions for each component object, which are called candidate positions. To decide the final answer from them, it is necessary to consider the candidate positions from all component objects together. Thus we construct a set of combination graphs based on all



(a) (b)

Fig 2-3: Illustration of matching the spatial relations using the pelvis model
(a) The semantic graph of the pelvis model
(b) The spatial relation constraint in the pelvis model

candidate positions. Since the semantic graph attribute stores the spatial relations among all component objects defined in the knowledge model, we will use this attribute to evaluate the combination graphs and reject those graphs with the high degree of distortion. Here, a flexible spatial relationship matching algorithm is employed. To explicate the match procedure, we use the spatial relation of pelvis model as an example. In Fig. 2-3, each dot represents the position of a component object. The whole algorithm is explained below:

Semantic Graph Match Algorithm

1. *Relax the semantic graph constraint by defining a similarity area which is the region enclosed by a circle around the location of each component object, as shown in Fig 2-3.(b)*
2. *Select one combination graph and match its candidate positions with the related component objects in the semantic graph.*
3. *If all component objects in the semantic graph have found their corresponding candidate positions in the combination graph, go to step 4, otherwise reject the combination graph.*
4. *For each component object, check its corresponding candidate position in the combination graph.*
5. If one of candidate positions lies outside of the related similarity area in the relaxed semantic graph, reject the combination graph, otherwise keep it.

Inference Engine

As mentioned in the previous section, image understanding is an inherently uncertain activity. In this research, we do not expect to get a perfect shape contour or an exact position of the ambiguous object at the first detection step. Instead, we prefer to keep some most possible candidates and desire to filter them after combining all component objects' candidates together, as it is more useful to get some imprecise results (maybe include several contours or positions but with at least one correct) than an uncertain result (only one contour or position but may not be correct). Thus, the inference engine must be capable of handling uncertainty at various levels of representation. Corresponding to the characteristics of these levels, different scoring techniques have to be adopted.

Scoring of Initial Segmentation Results

The problem of scoring the initial segmentation results is to determine a measure of the precision of the attributes from initial segmentation. As mentioned before, fuzzy logic approach is a fundamental and theoretically well-founded approach to measure the quality of results. It describes an attribute value into imprecise linguistic variables, fuzzy subsets in the fuzzy theory, and denotes an appropriate degree of membership that the attribute might belong to each fuzzy subset.

The definition of fuzzy set is a set of order pairs in a fuzzy space X

$$S = \{(x, \mu_s(x) \mid x \in X)\} \qquad (2.4)$$

`

$$Sup_{x \in X} \ \mu_s(x) = 1 \qquad (2.5)$$

Where $\mu_s(x)$ represents the degree of membership of *x* in *S*. Fig 2-4 shows the fuzzy sets and their membership functions for the gray level attribute. The domain of the fuzzy set is depicted along the x axis and ranges from black to white(0-255). The degree of

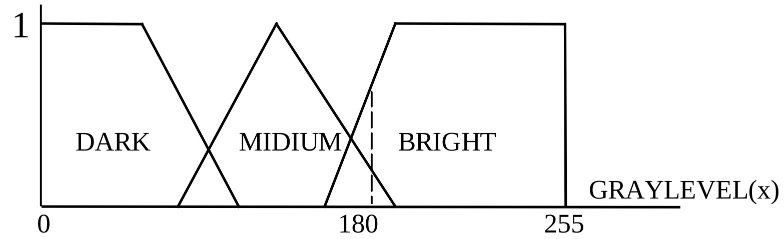


Fig 2-4: Fuzzy membership functions of the gray-level attribute

membership $\mu_s(x)$ can be seen along the vertical axis. the membership is between zero and one, zero representing no membership and one representing the complete membership. The fuzzy linguistic variables are labeled as DARK, MEDIUM, and BRIGHT. For instance, a gray level x=180 has 0.2 degree of membership in the MEDIUM fuzzy subset and 0.6 degree in the BRIGHT fuzzy subset. If the gray-level of expected object is BRIGHT, we can use these memberships and their corresponding fuzzy set to score the precision of the initial segmentation results.

Scoring of Object Recognition

On the recognition and understanding level, the encountered problem is how the judgment of instances of low-level processes can be combined to judgments of higher level processes. Values of attributes originated from initial segmentation, together with an estimate of their precision, should be evaluated by the requests of high-level constraints based on how well they fit the constraints. Again this may be done by fuzzy logic. First the constraints can be transformed into a set of fuzzy rules, then with the combination of the fuzzy score derived from initial segmentation and the fuzzy rules, a set of fuzzy solution can be generated. To make a final decision, a de-fuzzification process is employed and a numerical confidence will be assigned to evaluate each fuzzy solution, only the solution with high values are selected as the final results.

2.5 Implementation of the Object-oriented Knowledge Model

So far, the various aspects of the object-oriented knowledge model have been discussed and a corresponding model has been developed. The focus of this section will be on the application of this model into medical image analysis.

2.5.1. Chest Digital Radiographs

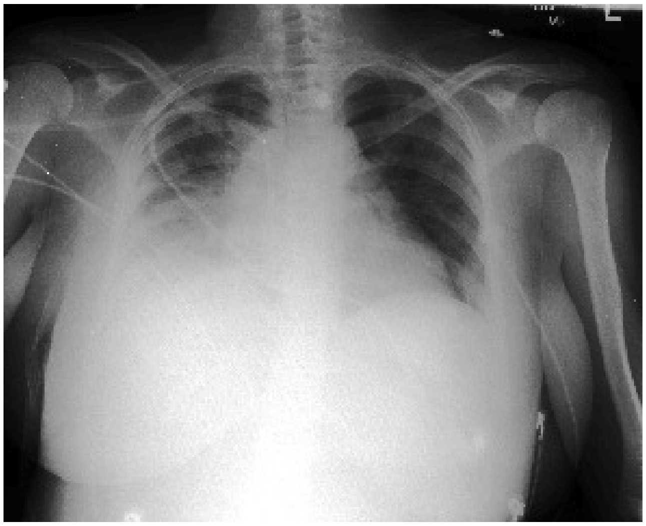
The utility of image processing technique in diagnostic chest radiology has increased with the growing acceptance of digital radiography. In this case study, we will incorporate an object-oriented chest model into the chest digital radiographs analysis. The major advantage of our model-based analysis is its robust property which works on both normal and abnormal radiographs.

When visually analyzing a digital chest radiograph, we use the following models to describe the contents enclosed in chest radiographs

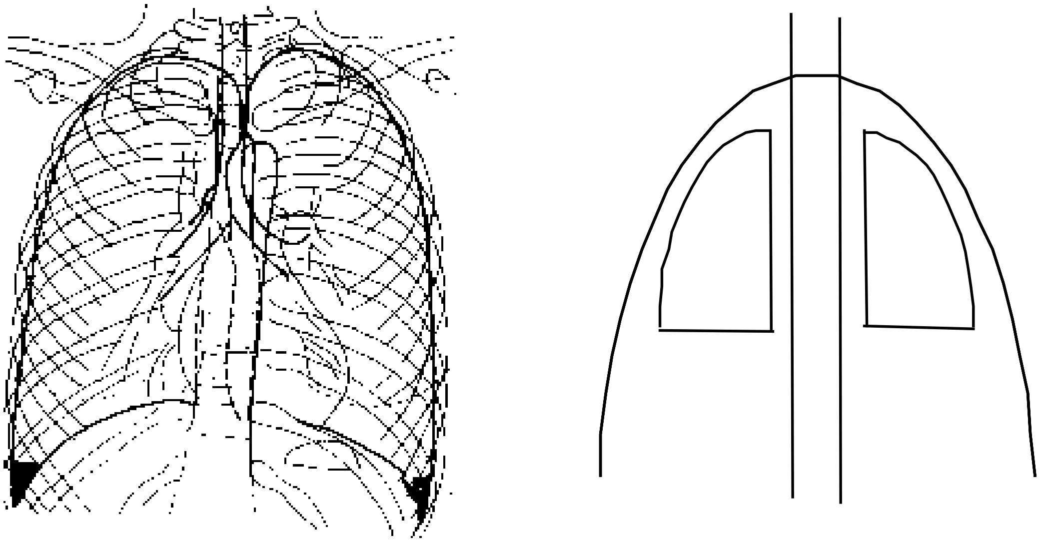
Model Design

Chest Model

Chest radiographs have complex anatomical structures as shown in Fig 2-5 (a-b). To simply the problem, we only focus on the major constitution anatomical structures in chest images as shown in Fig 2-5(c), which can be divided into three component object models: the body model, the spine model and the lung model. All these models are included in the main object class: the chest model class. Their semantic relations are



(a) The original chest radiograph



(b) The anatomical structure

(c) The chest model

Fig 2-5: Illustration of designing the chest model

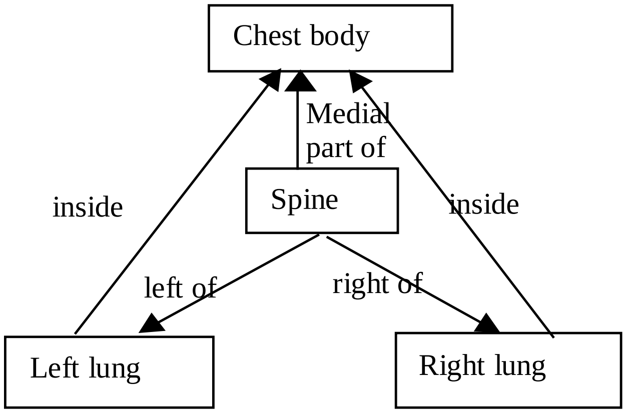


Fig 2-6 the semantic networks of the chest model

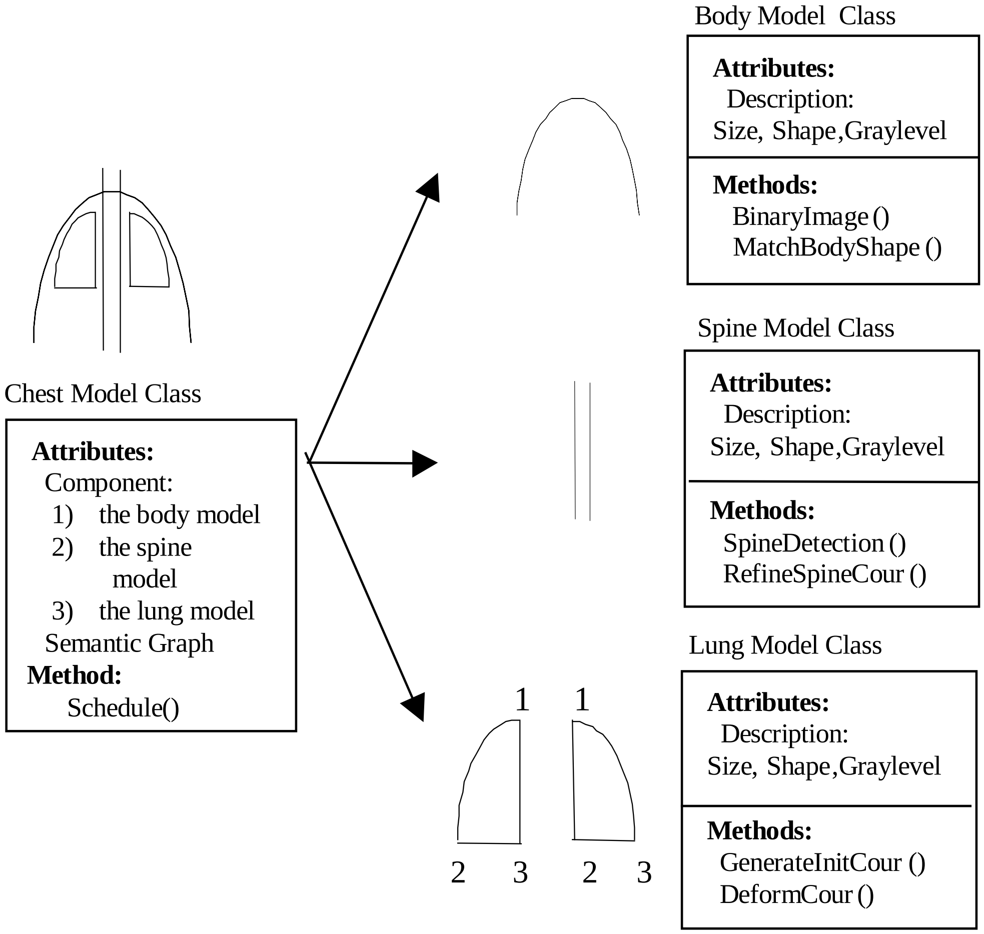


Fig 2-7: The class diagrams of the chest Model

depicted in Fig 2-6, and Fig2-7 demonstrates the class diagrams of all model objects.

Body Model

The body model is a major component object in chest radiographs. The detection of body model in chest radiographs allows us to obtain the size information of objects, which will be later used to scale model size. To detect the body part from chest radiographs, we employ the entropy thresholding technique (introduced in section 4.3.1) to automatically compute a threshold from the histogram of a radiograph, and then segment the radiograph with this threshold. The result binary image contains speckle and misclassified regions. This 'noise' can be easily removed by using opening operation. After binarization, the body model shape can be easily matched to the binary chest image.

Spine Model

Because of the symmetry of the human body, the spine can be served as the symmetrical axis of the chest body. To determine its location is useful to find out the orientation of the chest body and adjust the model orientation. With the help of the matched body model, we can roughly define the spine as the middle line of body model, then a fully automatic spine boundary detection algorithm is employed to extract true boundary of the spine column[Luo 99]. The algorithm consists of three steps: the first step detects the two starting points of spine edge. The second step integrates both region growing and edge tracing to generate initial contours of the spine. Using this initial contour, the final step employs a snake model algorithm to refine the spine contours.

Lung Model

The lung regions are the most concerned regions in chest radiographs analysis. Various image processing approaches have been applied to segment the lung regions in chest radiographs. To design an efficient model for lung region identification, the following knowledge is utilized in the lung model:

- Texture Characteristics: the visual appearance of lung and non-lung regions can be defined by the types of gray level pattern. For instance, lung regions are typically darker than non-lung regions and usually possess lower gray level values. Additionally, non-lung regions typically appear more flat and uniform than lung regions.
- Spatial Characteristics: This information can be used to predict locations in images where lung regions are most likely to occur than non-lung regions. For example, lung regions are approximately symmetric about the spine column and impossibly extend outside of body edges.
- Shape Characteristics: The anatomical shape of lung can be approximately modeled as lines and curves with the expected length, position, and orientation at the boundary points. Here, we define the shape of lung by using three boundary points as shown in Fig 2-7, where point 1 is the top of lung, point2 and 3 specify the lower left and right of the lung, respectively.

The initial contours of the lung regions are generated by first locating three boundary points, and then connecting them with smooth curves. In this study, we start from point 1, since it can be easily anchored based on the anatomical knowledge. After matching the body and spine models, we obtain a crossing points *A* of the body model outline and the spine center line. Then the position of point 1 can be derived as shown in Fig 2-8 by giving –x and y extensions on the horizontal and vertical directions, respectively. Once point 1 has been located in the image, we specify two lines starting from point 1 with 10^o and 45^o degree from the vertical direction. The first points on the lines which gray level changes from black to white in the binary chest image is defined as the boundary points 2 and 3, respectively. If it is difficult to find such boundary points, a pre-defined search length is used to locate them. Finally, we use two straight lines to connect point 1 and 2, point 2 and 3, and a polynomial curve to connect point 1 and 3. Once the initial contour is obtained, a robust snake model will be used to refine the contour.

Scheduling

Through our numerous experiments, we observe that it is easy to segment the body part in chest images, then with the assistance of body location, we can estimate the spine region and later derive the lung regions. Thus, the scheduling strategy of lung detection is designed as:

The body model match➔the spine model match➔the lung model match

Experimental Results

The performance of the proposed method was evaluated by 90 chest radiographs. The original radiograph set consisted of 2500 by 2048 pixels digital radiographs with 12-bit precision. In order to speed up the computation process, the original images were smoothed and subsampled to 625 by 512 pixels and the bit depth was reduced to 8-bit depth. The images shown in Fig 2-8 and 2-9 demonstrate the results of the proposed approach. Our experiments show that the proposed chest model works for nearly 80% chest images in our medical image database, and most results are similar to the demonstrated results.

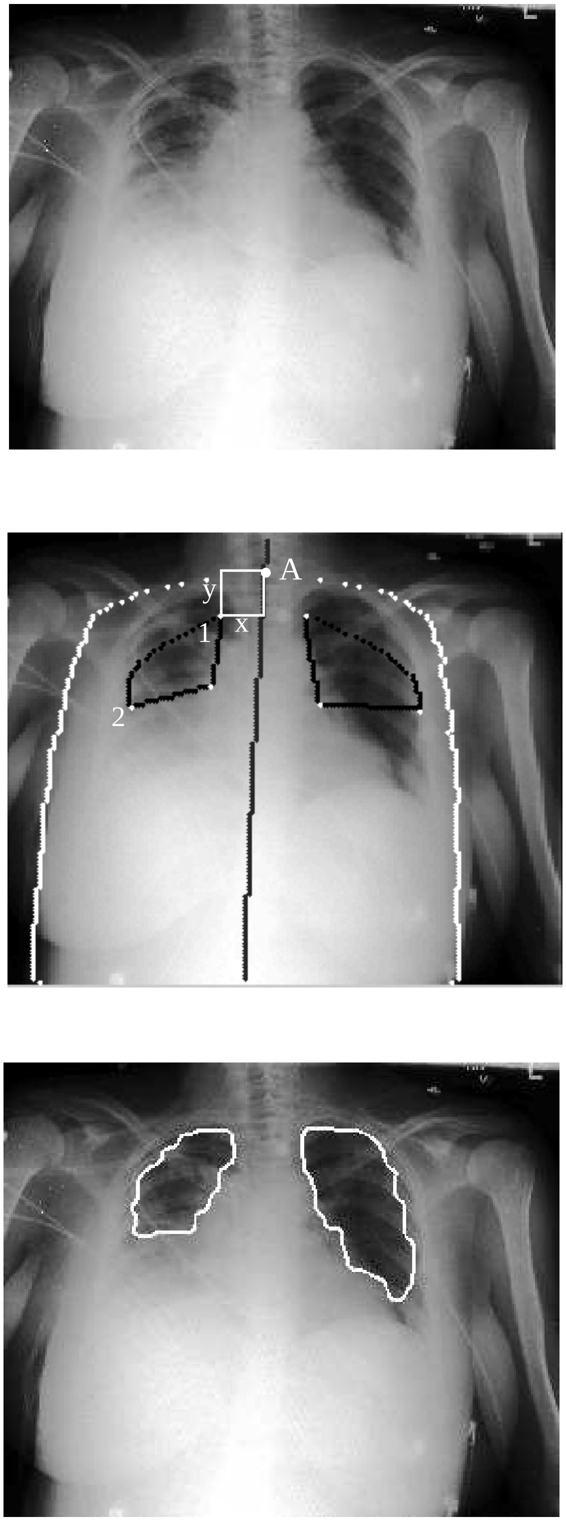


Fig 2-8: The experimental result of a chest radiograph

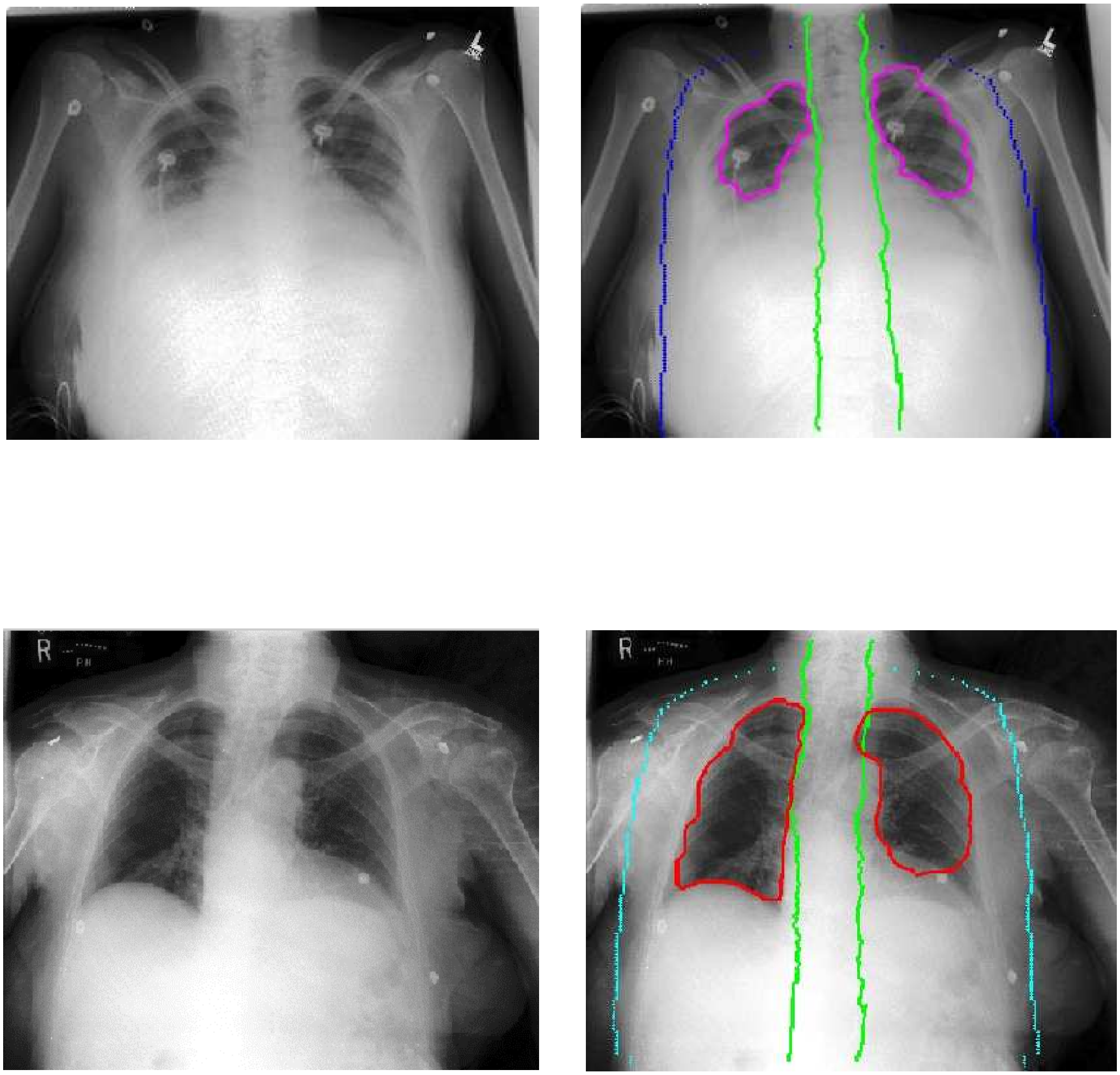


Fig 2-9: More experimental results of chest radiographs

2.5.2. Pelvis Digital Radiographs

This case study will use the object-oriented knowledge model to detect the major bone structures in pelvis images. For clearness, we still divide the model into two parts: the pelvis model attribute part and its method part.

The pelvis model attribute design

The pelvis images have a complex bone structure as shown in Fig 2-10(a-b). To simplify the problem, we focus on the major constitution bone structures in pelvis images as shown in Fig 2-10(c), which can be divided into four component object models: the hip model, the iliac bone model, the pubic bone model and the femur bone model. All these models are included in the main object class: the pelvis model class. Fig 2-11 demonstrates the class diagrams of all model objects.

The pelvis method design

To achieve a reliable bone contour detection in pelvis images, a well-defined control strategy is crucial. Through our numerous experiments, we observe some facts which exists in most pelvis images.

- The hip region is the easier part to be located in pelvis images
- The femur bones usually have more complete and clearer shape contour than other bone structures in the images.
- The pubic bone is located near the center of the hip region

Based on the information, the control strategy of the pelvis model is designed in the following procedure:

The hip model match ➔ the femur model match ➔ the pubic model match

➔ the iliac model match

In this case, the component objects do not have control strategies.

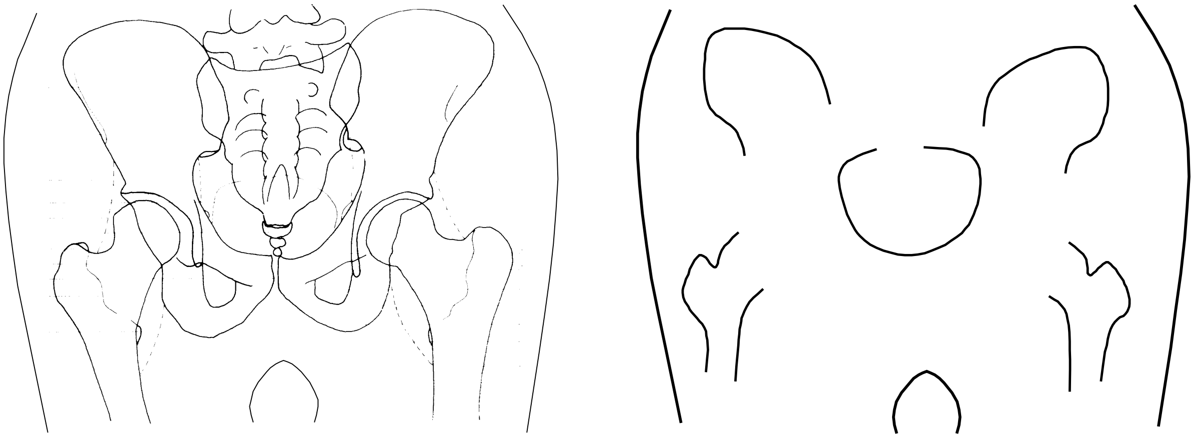
Experimental Results

The experiment has been performed on 50 pelvis images in our image databases. The images shown in Fig.2-12 and 2-13 demonstrate some experimental results. Table 2-1 lists the number of candidates and their corresponding confidence level of these images. From the original pelvis radiographs, it is hard to distinguish some bone contours even by human eyes, but our pelvis model can locate them pretty well.

Image No:	The candidate number	Confidence level	
Image 1	2	Candidate 1	2.345
		Candidate 2	2.158
Image 2	2	Candidate 1	2.745
		Candidate 2	2.019
Image 3	5	Candidate 1	3.012
		Candidate 2	2.754
		Candidate 3	2.314
		Candidate 4	2.109
		Candidate 5	1.982
Image 4	2	Candidate 1	2.189
		Candidate 2	1.974

Table 2-1: The candidate number and their corresponding confidence level of pelvis radiographs

(a) The pelvis radiograph



(b) The pelvis anatomical structure

(c) The pelvis model

Fig 2-10: Illustration of designing of the pelvis model

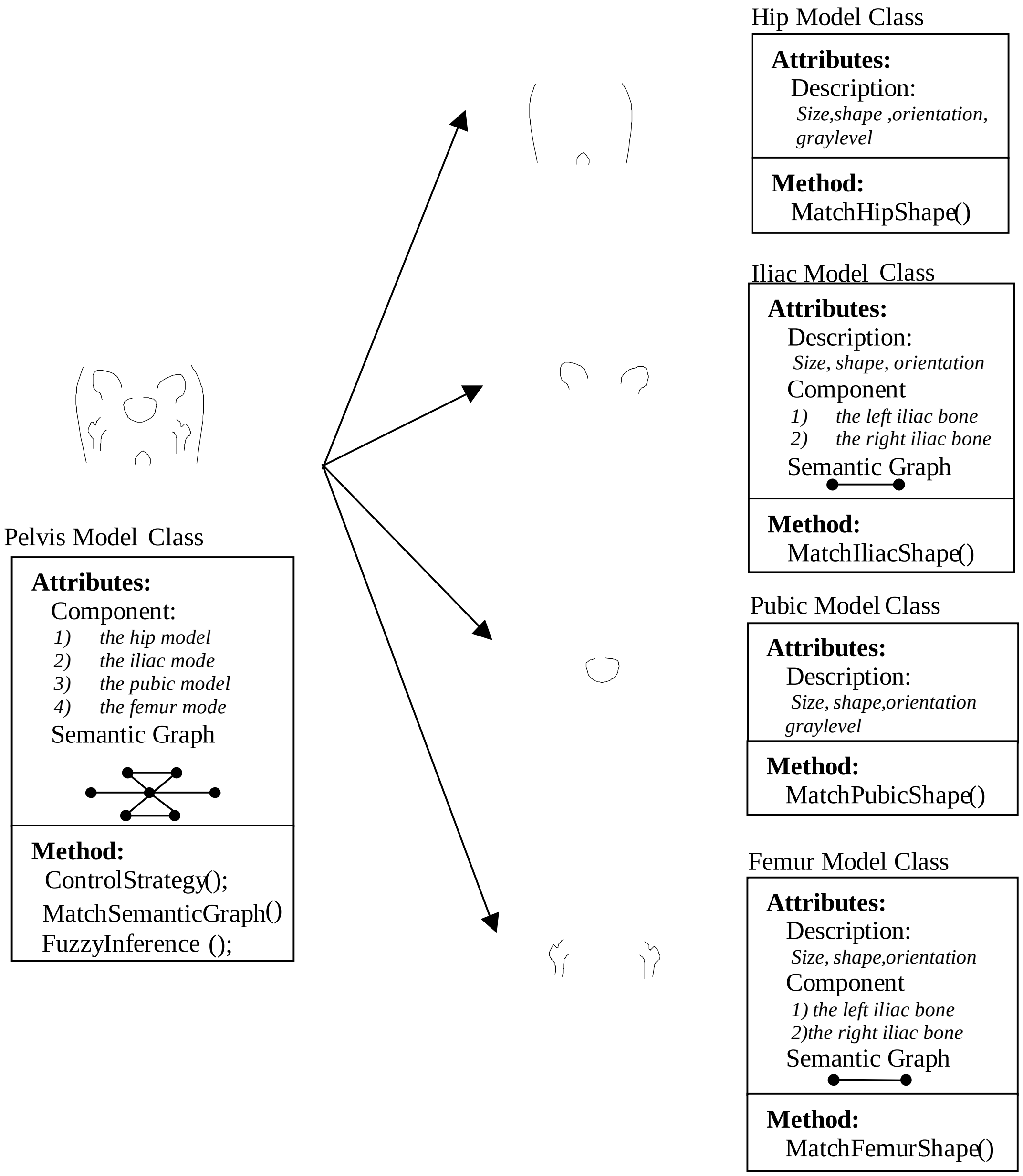
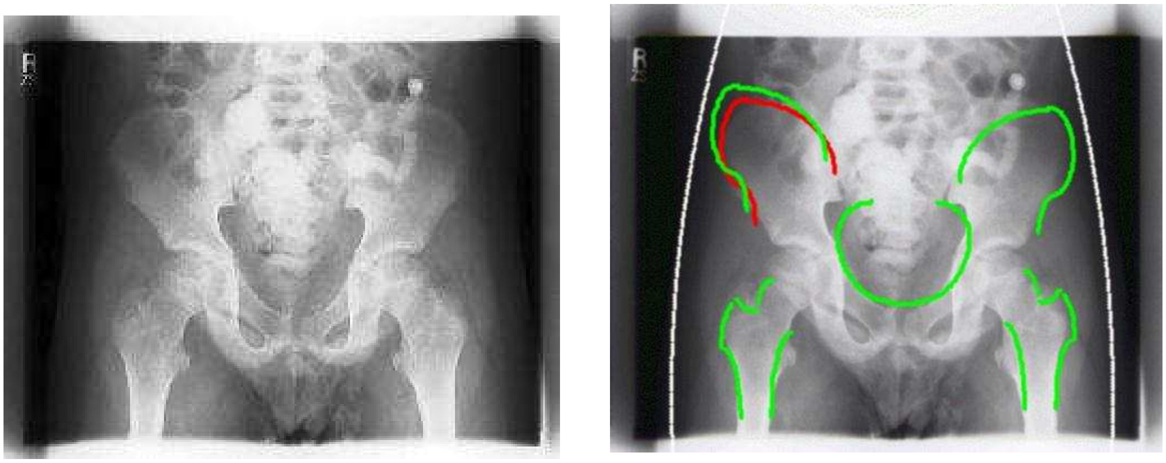
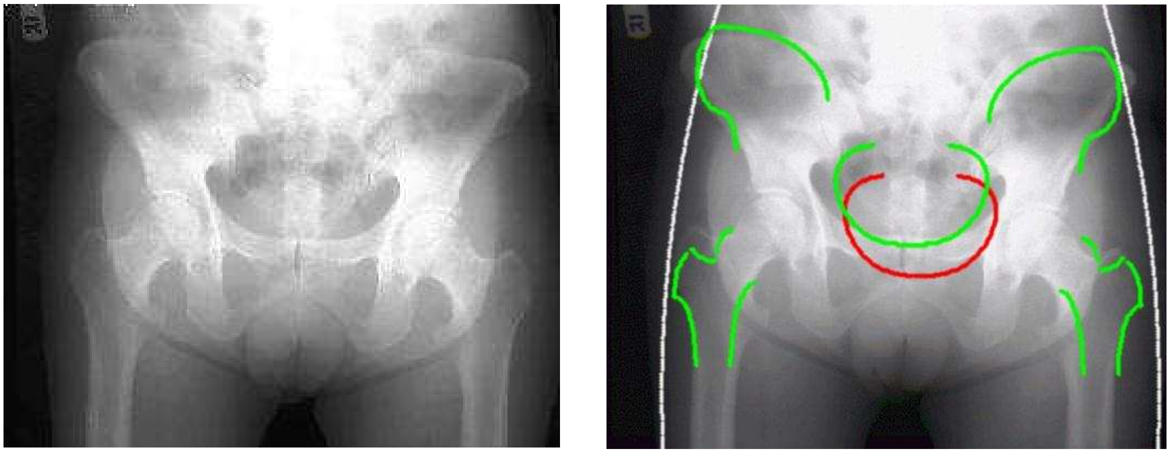


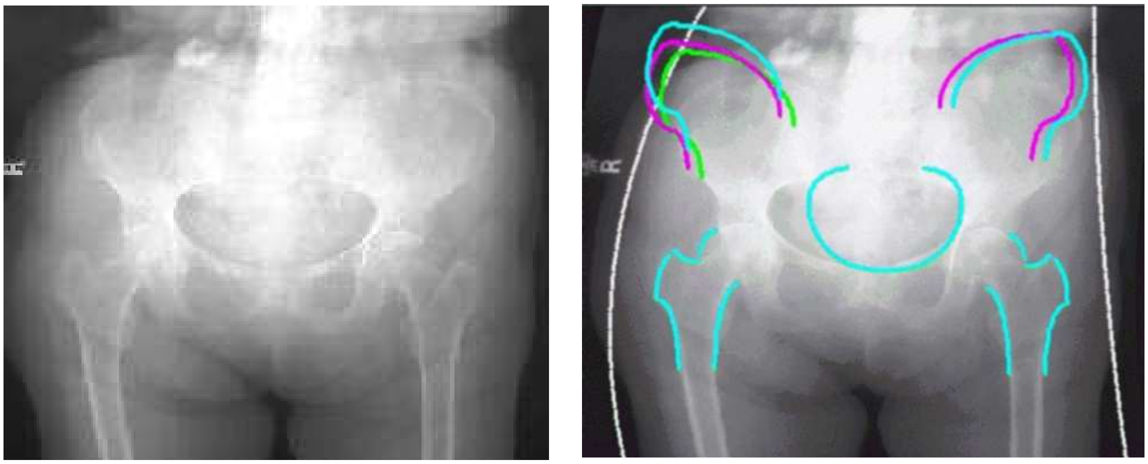
Fig 2-11: The class diagram of the pelvis model



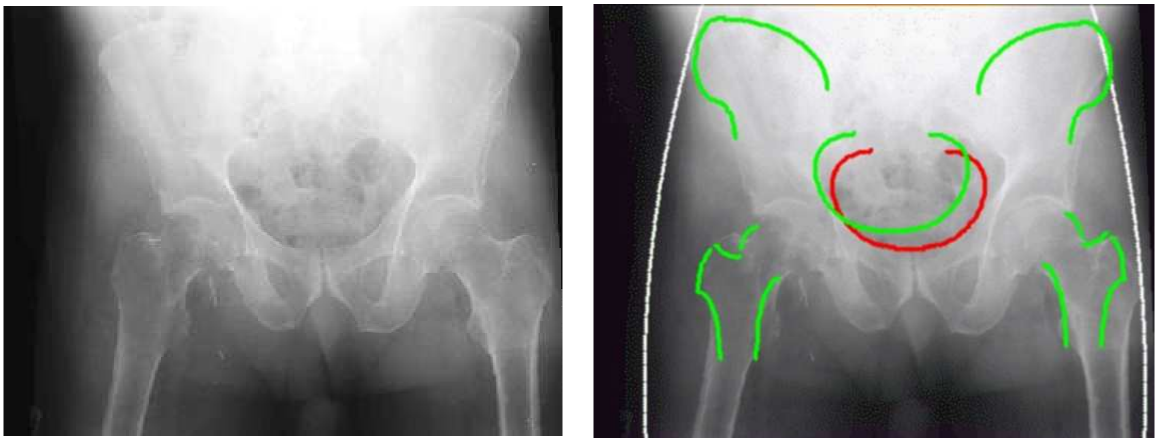
(a) Image 1



(b) Image 2



(c)   Image 3

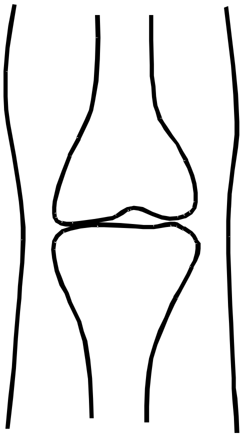


(d)   Image 4

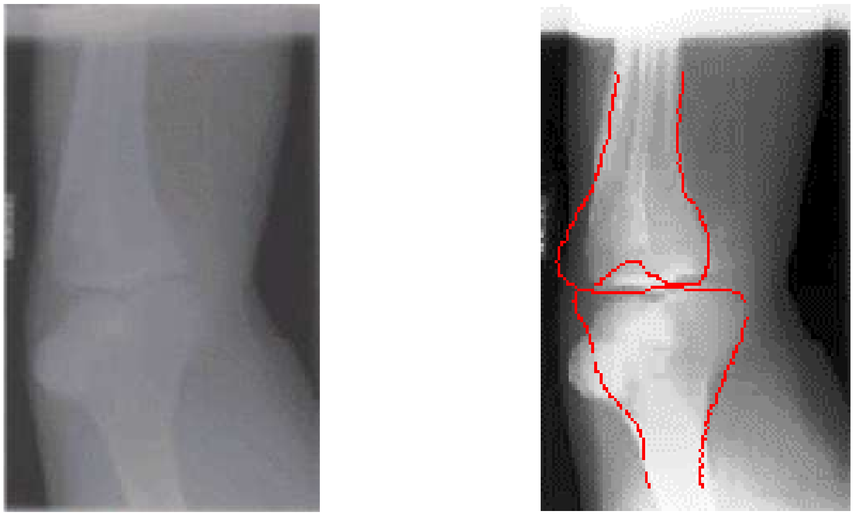
Fig 2-12: The experimental results of pelvis radiographs

2.5.3 Other Digital Radiographs

To test the robustness of the proposed object-oriented knowledge model, we also applied the model to different body parts in our image databases, which contains 49 knee images, 64 ankle images, 35 elbow images, The following figures demonstrate some experimental results, which can represent most processing results by using the defined models.

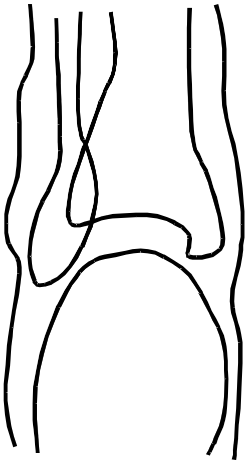


(a) The knee model



(b) The original image (c) The experimental

Fig 2-13: The experimental results of knee radiographs



(a) The ankle model

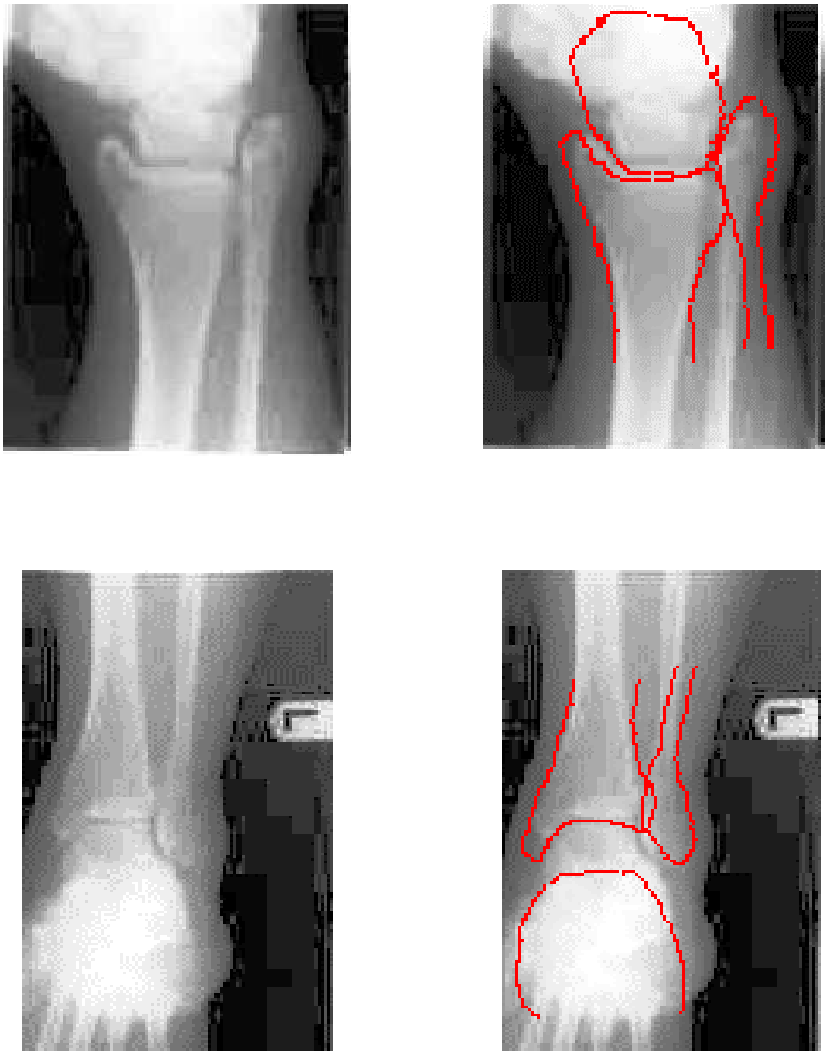
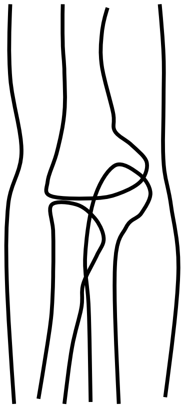


Fig 2-15: The experimental results of radiographs



(a) The elbow model

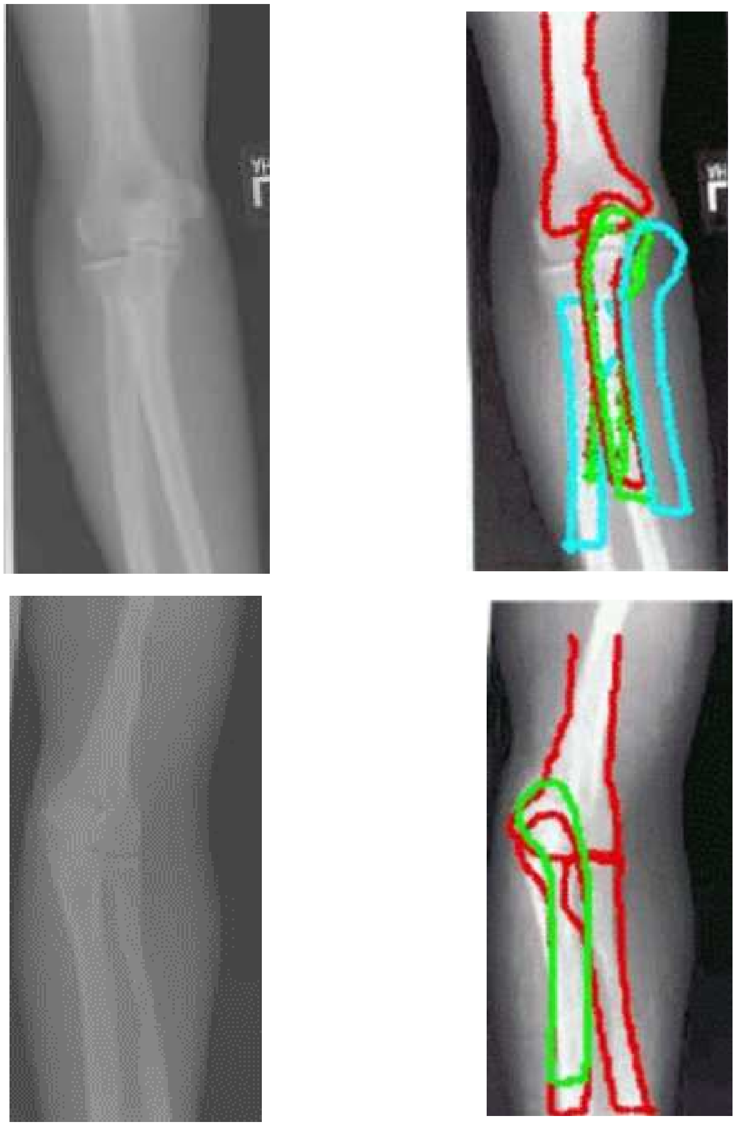


Fig 2-15: The experimental results of elbow radiographs

2.6 Summary

In this chapter, we proposed a new knowledge model for medical image content analysis based on the object-oriented paradigm. The new model aims to abstract common properties from different types of medical images, but is able to be extended to a particular type image. To achieve this, we design the model content into two main parts: the attribute part and the method part. The attribute part describes characteristics of the model by three attributes: the description, component, and semantic graph attributes. The method part specifies some actions conducted by the model, such as scheduling the detection procedure, properly deforming the model component shapes to match the corresponding anatomies in images, selecting the best match candidates, and verifying the combination graph from detected candidates with the semantic graphs defined in the model. The performance of the proposed model has been tested on various medical images, such as chest, pelvis, elbow, ankle, and knee digital radiographs. Initial experimental results are encouraging.

Chapter 3:

Robust Snake Model

3.1 Introduction

The segmentation of a given image is one of the most important techniques for image analysis, understanding, and interpretation. Generally speaking, Feature-based image segmentation includes two basic image processing techniques: the boundary-based (or edge-based) segmentation and the region-based segmentation.

Under the Marr’s paradigm [Marr 80], boundary extraction has conventionally been treated as two independent problems: edge detection and edge linking. At the lowest level, an edge detector operates directly on the image intensities to produce an edge map marking the location, strength and direction of edge points. An edge linking algorithm then yields an ordering of successive edge points based on some predefined criterion functions such as continuity and connectivity. The enormous amount of research in edge detection can be broadly classified into one of several categories: optimal filtering [Canny 86], scale space [Marr 80], surface fitting [Haralick 84] and cost minimization [Tan 92]. A great deal of literatures also exists on edge linking [Ashkar 78, Marteli 72].

Despite this tremendous effort, there is little consensus that the problem of boundary extraction has been solved. A problem with the boundary-based approach is that errors made in edge detection propagate to edge linking without opportunity for correction. Moreover, due to image noise and the image projection process, the lower level visual tasks are generally ill-posed [Poggio 84], i.e., they are under-constrained and do not have unique solutions in general.

Active contour, also known as snake or deformable model, was originally proposed by Kass et. al [Kass 87] as a regularization approach to the ill-posed edge detection problem and has been proven to be an effective method in contour extraction and image interpretation. The snake model advocates the paradigm that the presence of an edge depends not only on the gradient at a specific point but also on the spatial distribution of the gradient. The snake incorporates this global view of edge detection, by assessing continuity and curvature combined with the local edge strength, to determine an edge. The principal advantage of this method over other feature extraction techniques is that it integrates image data, an initial estimate, desired contour properties and knowledge-based constraints into one single extraction process, along with the provision of an implicit description of the boundary.

This chapter discusses the significant properties of the snake model and the problems of implementing the model to provide a basis for designing a robust snake model. Section 3.1 introduces the basic concepts of snake model. Its major properties, such as energy formulation, minimization technique, initialization, are investigated in section 3.2. With the motivation of improving the behavior of the traditional snake model, section 3.3 introduces a new region model which directly combines the region and edge information into segmentation. Based on the framework of this model, we reformulate both the internal and external energies of the traditional snake model, and the force function of level set model to make snake models more robust. Section 3.4 demonstrates the performance of the proposed robust snake models, followed by the conclusion in section 3.5.

3.2 Snake Model

The snake model is built on an energy minimization framework. A contour has two energy, or cost, functions associated with it: an internal and an external energy. The internal energy measures desired properties of a contour's shape, such as smoothness and continuity. The external energy is derived from a selected image, to measure desired image features such as edges or lines. The technique refines an estimated initial contour by an energy minimization technique. Under this energy minimization framework, the internal energy, the *a priori* information, regularizes the external energy, the *a posteriori* information.

A snake ***s*** can be represented an ordered set of points, $\boldsymbol{V}=[\mathbf{v}_1, \mathbf{v}_2 \ldots \mathbf{v}_n]$ where each snaxel $\mathbf{v}_i$ is defined on the finite grid:

$$\mathbf{v}_i(s)=(x_i(s),y_i(s)). \qquad i=1\ldots n$$

A snake can be either closed or open, depending on whether the end points are connected.

The total energy of a snake model can be expressed as follows:

$$\begin{aligned} E_{snake} &= \int_0^1 E_{snake}(v(s))ds \\ &= \int_0^1 E_{\text{int}}(v(s)) + E_{img}(v(s)) + E_{cont}(v(s)))ds \end{aligned} \qquad (3\text{-}1)$$

where: E_{int} represents the internal energy of the spline due to bending

E_{img} gives rise to the image forces

E_{cont} denotes the external constraint forces

3.2.1 Internal Energy

The internal energy E_{int} of a snake is formed from Tikhonov stabilizer [Tikhonov 63], and is defined to be

$$E_{\text{int}} = (\alpha(s)\left|\frac{dv(s)}{ds}\right|^2 + \beta(s)\left|\frac{d^2v(s)}{ds^2}\right|^2)/2 \tag{3-2}$$

The first order continuity term, weighted by $\alpha(s)$, makes the snakes act like a membrane and the second order curvature term, weighted by $\beta(s)$, makes the snake act like a thin plate. Adjusting the weights $\alpha(s)$ and $\beta(s)$ controls the relative importance of the membrane term and the thin plate term. Setting $\beta(s) = 0$ at a point $\boldsymbol{s}$ allows the snake become second-order discontinuous at that point and develop a corner. Setting $\alpha(s) = \beta(s) = 0$ at a point s allows the snake to become discontinuous.

It should be noted that $\alpha(s)$ and $\beta(s)$ are defined to be functions of curve parameter, and hence segments of the snake may have different natural behaviors. This property can only be exploited when high-level information about the desired contour is known or can be estimated. Minimizing the energy of derivatives intuitively gives a smooth function, and is equivalent to filtering the contour by a low-pass filter.

3.2.2 Image Energy

In order to attract snakes to salient features in images, the external energy is needed. The typical external energy is derived form the image data over which the snake lies. Kass et. al. [Kass 87] presented three different energy functions that attract a snake to lines, edges, and terminations. The total energy can be expressed as a weighted combination of these three different functional.

$$E_{image} = w_{line}\,E_{line} + w_{edge}E_{edge} + w_{term}E_{term} \qquad (3\text{-}3)$$

Line Functional

The line-based functional is very simple

$$E_{line} = I(x,y) \qquad (3\text{-}4)$$

Where $I(x,y)$ denotes the image intensity at an image point (x,y). The sign of w_{line} specifies whether the snake is attracted to light or dark lines.

Edge Functional

A suitable edge functional is

$$E_{edge}{}^{1}(x,y) = -|\nabla I(x,y)|^{2} \qquad (3\text{-}5)$$

$$E_{edge}{}^{2}(x,y) = -|\nabla [G_{\sigma}(x,y *I(x,y))|^{2} \qquad (3\text{-}6)$$

Where $G_{\sigma}(x,y)$ is a two-dimensional gaussian function with standard deviation σ, and ∇ is the gradient operator. This functional attracts the contour to points with high gradient.

Termination Functional

Line terminations and corners may influence the snake using a weighted energy functional E_{term}, Let $\boldsymbol{g}$ be a slightly smoothed version of the image $\boldsymbol{I}$, let $\psi(x,y)$ denote the gradient directions along the snake in the smoothed image $\boldsymbol{g}$ and let

$$\mathbf{n}(x,y) = (\cos \psi(x,y), \sin \psi(x,y)) \qquad (3\text{-}7)$$

$$\mathbf{n}_{R}(x,y)=(-\sin \psi(x,y), \cos \psi(x,y)) \qquad (3\text{-}8)$$

be unit vectors along the perpendicular to the gradient directions $\psi(x,y)$. Then the curvature of constant-gray-level contours in the smoothed image can be written as

$$E_{term} = \frac{\partial \psi}{\partial n_R} = \frac{\partial^2 g / \partial n_R^2}{\partial g / \partial n} \qquad (3-9)$$

$$= \frac{(\partial^2 g / \partial y^2)(\partial g / \partial x)^2 - 2(\partial^2 g / \partial x \partial y)(\partial g / \partial x)(\partial g / \partial y) + (\partial^2 g / \partial x^2)(\partial g / \partial y)^2}{((\partial g / \partial x)^2 + (\partial g / \partial y)^2)^{3/2}}$$

The snake behavior may be controlled by adjusting the weights w_{line} , w_{edge} ,w_{term}.

3.2.3 External Constraint Energy

The constraint energy comes form external constraints imposed either by a user or some other higher-level processes, which may force the snake toward or away from particular features. If the snake is close to some desirable features, the energy minimization will pull the snake to the desired solutions. However, if the snake settles in a local energy minimum, and a higher-level process determines it as incorrect, then an external constraint may be made at this location to force the snake away from the local minimum.

3.3 Properties of Snake Model

The snake model provides a method of integrating boundary extraction within an energy minimization procedure, which is a convenient way to incorporate many constraints. It defines solutions to be the minima of a contour measuring energy function. The choice of this energy function is a compromise between integration of a priori generic shape constraints and image data.

In this section, the properties of snake techniques are investigated in detail, and we will forces on discussing energy formulation, initialization, and implementation with respect to the energy minimization framework of the snake model. The motivation of discussion is to determine the emphasis for developing new improved techniques.

3.3.1 Energy Formulation

The energy function has two components (In most cases, the image energy and the external constraint energy are combined as the external energy). The internal energy function measures the shape of the contour and the external energy function measures the boundary information derived from the image data. The internal energy, defined in Equation 3-2, measures the energy in the first and second derivatives of the contour, with respect to normalization arc length. This shape measure is a soft constraint since it is not enforced, but it pre-disposes the solutions to continuous and smooth contours.

A weakness of the generic constraints of the internal energy function is that they do not easily accommodate local variation in the shape of the contour. Kass et.al. [Kass 87] proposed the internal weighting parameters α and β to be functions of the contour parameter *s*, While Williams et. al.[Williams 92] showed that setting β to zero at a selected point allows the contour to become second discontinuous and develops a corner there. Generally speaking, it is difficult to relate these parameters explicitly to the desired shape of the contour, which results in the unpredictability of the internal energy. Usually, snaxels with high curvature may have a much stronger internal energy than that of snaxels with low curvature.

In practical applications, the snake model is implemented on the discrete grid. So the internal energy E_{int} can be approximated as

$$E_{\text{int}}(v_i) = \frac{\alpha|v_i - v_{i-1}|^2}{2h^2} + \frac{\beta|v_{i-1} - 2v_i + v_{i+1}|^2}{2h^4} \qquad (3-10)$$

where h is the space difference between the snaxels. This discrete approximation introduces several problems. Firstly, E_{int} vanishes only when all snaxels have the same position, which often causing a snake to shrink around strong edge points. Since the primary purpose of E_{int} is to impose smoothness constraint on snakes. Therefore, in the absence of external force, E_{int} should force the snake to take the form of a *line* for an *open* snake, or a *circle* for a *closed* snake. Thus, to counter the shrinkage tendency, ballooning force [Cohen 91] and hard constraint on snaxel proximity [Fua 97] have been suggested, but they required prior information in setting the weighing factor or constraint. Secondly, E_{int} in Equation 3-10 is not invariant to scale, rotation and translation changes. One of the challenges with the traditional snakes and the balloon snake is the difficulty of adjusting the space between the nodes as the curve evolves. When a snake expands, the resolution of the snake is reduced if the number of node points stays constant. During contraction, the snaxels move much closer to each other, which results in unnecessary computational expense. Therefore, one must dynamically adjust the positions of the snaxels when the snake undergoes large changes.

Regarding to the external energy, most snake models only use the edge information (usually the gradient feature of images) as the external force. Since the magnitude of the edge-based external energy dies out quite rapidly away from the object boundary. So if a contour is far away from the object boundary, there will be no external force acting on it. Even near boundaries, if the external force is generated by a weak edge, the contour is likely to overrun the edge and continue its progress. To attack both problems, Cohen, et al [Cohen 91] proposed an external force model that originates from an edge map of images and can provide a large capture range. Recently, Xu et al [Xu 98] proposed a new external force called gradient vector flow (GVF) which is computed as a diffusion of the gradient vector of a gray-level or binary edge map derived from images. The particular advantage of the GVF snake over other snakes is its ability to move into boundary concavities. Instead of using the edge information as the external energy, some attentions have been given to the use of the interior region information. Ronfar [Ronfard 94] introduced an energy function based on statistical models of the background and object region. The strategy is to push or pull pixels on the contour if their neighbors fit the background or object model distribution. Ivins et al [Ivins 94] proposed a statistical snake which expands or contracts according to a pressure force linked to the image features. Cox et al [Cox 96] utilized a computationally efficient graphs partitioning algorithm which minimizes a ratio between the exterior "boundary cost" and interior "region benefit' to extract the best contour from images.

Another factor which directly affects the solution is how to select the proper weights for the internal energy and external energy terms. The internal energy imposes continuity and smoothness constraint, and the external energy attracts the snakes to salient image features. If the internal energy overweighs the external energy, the snake model will yield smoothed curves which are robust to noise but insensitive to image features. In contrast, if the external energy has a major influence, it enables the snakes to effectively capture boundary discontinuities, but it also makes the solution very noise sensitive. Despite its significance, this problem is either ignored [Kass 87], or based on empirical observations [Amini 90, Cohen 91] and heuristics[Williams 92], Lai [Lai 94] proposed a local minmax criterion which automatically determines the optimal tradeoff at every location along the boundary of interest without any user input or heuristics. Although this method improves the performance of energy function, it still suffers the weights problem introduced by the original definition.

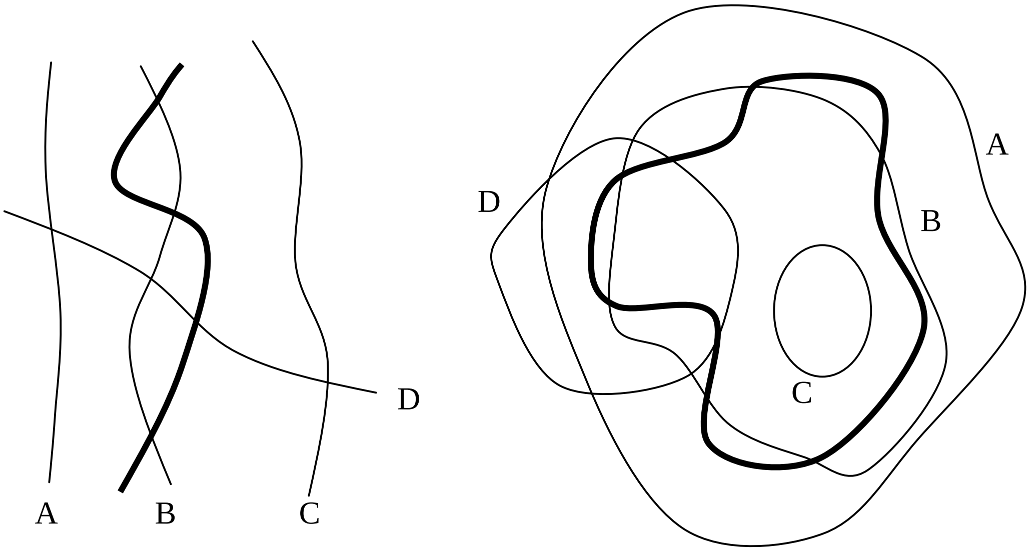
3.3.2 Initialization

Snakes can be initialized by using user interaction, influence from a higher-level process, or information from image data adjacent in time or space. These interactions provide an approximate starting position and shape for the snake near the desired contour. The topology of the initialization with respect to the desired contour may take many forms. A representative collection is illustrated in Figure 3-1 for both open and closed contours. In this figure, the bold lines represent the desired boundaries of objects. These initializations may be grouped as follows:

(A) outside
(B) overlapping
(C) inside
(D) perpendicular

where contours A,B and C are well-aligned with the solution and contour D is badly aligned with the solution.

The original technique of Kass et. al. [Kass 87], in the absence of any image forces, contracts the contour under the influence of its internal forces. An open contour will shrink to a point unless the ends are fixed. Similarly, a closed contour will contract to a point in the absence of any image forces. Accordingly, the open contour requires an initialization which is within the valley of the image energy, such as Class B in Figure 3-1(a), in order to provide an appropriate solution. The closed contour will contract and find a solution for convex contours if the contour is external to it and unobstructed, like class A in Figure 3-1(b). For non-convex contours, the situation is further complicated by the



(a) An open contour (b) A closed contour

Figure 3-1: Initial contour configurations for open and closed contours

increased dependency on the internal parameters. Both non-convex and convex contours will only work correctly with a subset of class B.

Kass proposed a ‘volcano’ force and additional spring forces to push or pull the contour towards image features, but the problem is that these forces are not integrated with the contour, and consequently their strength and position have to be determined manually, leading to a highly interactive approach. To tackle the initialization, the following issues should be considered.

The contraction force generated from the internal energy is not only an unsatisfactory force to work on the snake contour, but also compounds the problem of determining values for the parameters. Therefore, it is inappropriate for the snake model. If the initialization is inside of the feature, contraction cannot drive the contour towards the solution. Various methods have been proposed to address this problem. Cohen [Cohen 91] proposed a method using an additional normal force to inflate the contour. The normal-driven method has particular advantages over the conventional snake in the extraction of non-convex boundaries. The force is implemented by adding F_{normal}, defined as:

$$F_{normal} = \rho \boldsymbol{n}(s) \quad (3\text{-}11)$$

where ρ is the parameter determining the strength of the normal force. $\boldsymbol{n}(s)$ is a unit vector normal to the contour curve. Depending on the sign of ρ, the contour can be made to inflate or deflate. Sullivan et. al. [Sullivan 93] developed a similar pressure force that acts along the line joining the center of gravity to the edge of the contour. However, in all these cases, the contraction force still remains and complicates the determination of the normal force parameter. Furthermore, the normal driven techniques allow greater flexibility but include additional parameters. Due to the fact that the contraction force depends not only on the internal parameters but also on the contour shape, Xu et al [Xu 94] proposed to add a new pressure force based on the internal parameters to totally offset the shrinking force. As a result, their internal force is virtually zero, which makes the modified model lose the smoothness while deforming.

Another choice to relax the constraints of initialization is to enlarge the capture range of external energy. A simple and direct way to accomplish it is to increase σ of Gaussian function, but the larger σ will result in inaccurate boundary localization. Several methods such as distance potential force[Cohen 91], gradient flow vector force[Xu 98] have been proposed to significantly increase the capture range of the traditional snake. But they all use only edge information. In practical applications, edges in an image can be boundaries of the interested object and other objects in the image as well as noise. Therefore, the snake optimization function may have many local optima. As an opposite approach in image processing, region-based methods partition an image based on homogeneity of spatially localized feature. The advantage of region-based methods is that image information inside the object is considered, as well as that on its boundaries. Also it can tolerate certain noise. Some attentions have been given to the use of the statistical information from interior regions, like [Ronfard 94, Ivins 94, Cox 96]. Also, there is a significant effort to integrate boundary-based with region-based segmentation approaches [Chakraborty 96, Zhu 96, Pavlidis 90, Paragios 99]. However, the difficulty lies in the fact that even though these two approaches yield complementary information, they involve conflicting and incommensurate objectives. The region-based methods attempt to capitalize on homogeneity properties, whereas boundary-based methods use the non-homogeneity of the same data as a guide.

3.3.3 Energy Minimization

The choice of an energy minimization framework for active contour techniques is advantageous in that it is a convenient way of integrating many constraints. The term ‘energy’ derives from a mechanical analogy in which the snake energy terms have analogous potential energies; the active contour model defines contour solutions to be minima of the snake energy function Equation 3-1. A major concern is the location of these minima. Active contours differ from most conventional boundary extraction techniques in that they search for a local minimum as well as a global minimum.

Originally, a resolution minimization method proposed by Kass et. al. [Kass 87] estimated partial derivatives in *s* and *t* by the finite-differences method. Later, a dynamic programming approach [Amini 90] was proposed which allows ‘hard’ constraints to be added to the snake. Difficulties with the numerical instability of the original method were overcome by Berger [Berger 90] by incorporating an idea of snake growing. In this method a single primary snake may begin which later divides itself into pieces. The pieces of very low energy are allowed to grow in directions of their tangents, while higher-energy pieces are eliminated. At every growing step, the energy of each snake piece is minimized. This snake growing method may overcome the initialization problem by using the very-low-energy parts of the primary snake to initialize the snake growing in later steps. The iterative snake growing always converges and the numerical solution is therefore stable – the robustness of the method is paid for by an increase in the processing cost of the algorithm.

A different approach to the energy integral minimization that is based on a Galerkin solution of the finite-element method was proposed in Cohen[Cohen 93] and has the advantage of greater numerical stability and better efficiency. This approach is especially useful in the case of closed or nearly closed contours. An additional pressure force is added to the contour interior by considering the curve as a balloon which is inflated. This allows the snake overcome isolated energy valleys resulting from spurious edge points, giving better results all over. Another approach using a finite-element method [Karaolani 92] also significantly improves the solution's efficiency.

3.3.4 Implementation

The main advantage of the original implementation [Kass 87] was its implicit nature which allowed internal forces to be propagated around the contour in a single iteration. However it requires that any external energy, such as the image energy, was differentiable, and needed careful consideration in the choice of parameters to provide stable behavior. Many alternative implementation and extensions to the original technique have been proposed. Dynamic programming has been demonstrated as a universal technique, including a local minimum implementation [Amini 90] as well as its more familiar role as global minimum technique [Cohen 97]. An advantage is that it is compatible with the inclusion of hard constraints. A fast implementation proposed by [Williams 92] , the greedy algorithm, is based around a neighborhood search. The local nature of this approach however, implies poor performance in the presence of noise and does not guarantee a local minimum solution. Lai [Lai 94] modified it to include a guarantee but at a cost of a significant speed penalty. Davatzikos [Davatzikos 95] has shown that the closed contour problem can be solved within the Fourier domain. However the generalization to open contours is difficult due to the extra initial conditions. Tsai [Tsai 93] has configured a Hopfield network to implementation the minimization. The rest of this section focus on analyzing three important issues related to the implementation.

Contour representation

The original description of the contour is continuous in ***s***. However, in order to implement the snake model, it is convenient to convert the continuous snake model equation into a discrete model which defined as

$$E_{snake}(v) = \sum_{i=0}^{N-1} E_{\text{int}}(v_i) + E_{image}(v_i) \qquad (3-12)$$

The corresponding discrete equations for the internal energies are:

$$E_{cont}(v_i) = \frac{\alpha \left| v_i - v_{i-1} \right|^2}{2h^2} \qquad (3-13)$$

$$E_{curv}(v_i) = \frac{\beta \left| v_{i-1} - 2v_i + v_{i+1} \right|^2}{2h^4} \qquad (3-14)$$

Where E_{cont} and E_{curv} are the continuity and curvature energies respectively. These internal energies are required to promote even spacing of the contour points to regulate the contour behavior.

The implementation of the model introduces an approximation. Therefore a strategy for determining the number of points representing the contour, or the sampling rate is required. As we mentioned, this discrete approximation introduces the contraction force and gives unwanted results. Additionally, the sampling may introduce extra local minma within the energy function.

An alternative approach employs continuous contours which inherently provide a complete description of the contour, either by using finite elements [Cohen 93], or similarly by using local polygonal approximations [Neuenschwander 94], or the B-spline formulation of Menet et. al.[Menet 90]. The image energy can now be integrated along the length of the contour, rather than at discrete points. However this approach is computationally expensive.

Contour Space

There are two possible alternatives for the contour space. Either the contour can be restricted to take any position in a continuous space or it can lie on the pixel grid.

The advantages of using a continuous space for the snake points is that it allows the even points spacing to be effected, and allows the possibility of sub-pixel resolution, provided by the interpolation ability of the snake. To accomplish a continuous representation requires the value of the image function at any position in the image space. Discontinuities at pixel boundaries impose difficulties, but it can be solved by using a simple linear interpolation scheme.

Ivins [Ivins 94] argued that because the contour model is representing discrete data, it is only necessary to record the boundary to pixel accuracy. The advantage of a discrete space is that the whole search space is finite, which is very important in energy minimization techniques and search-based techniques. Another particular advantage is that convergence is easier to guarantee. However, Evolutionary techniques are not suited to a discrete based grid because of problems with introduction of extra local minima within the internal energy function. A discrete space can cause a contour to stop as a consequence of an imposed grid, when a continuous model would continue to contract or evolve.

Convergence

The stopping criterion is simpler for discrete search based techniques. A suitable criterion concerns a termination parameter δ, which measures the rate of a snake evolution at time t and $t+1$.

$$\max_i \left| v_i^{t+1} - v_i^t \right| < \delta \qquad (3-15)$$

Normally the rate of the snake evolution must remain within a neighborhood of one pixel in one iteration in order that all appropriate data is considered. This limits the speed of the evolutionary technique and is comparable with the size of the discrete space which controls the speed of search-based technique. In order to for an evolutionary approach to consider all appropriate image data, it must not advance more than one pixel in any one iteration.

3.4 Robust Snake Model

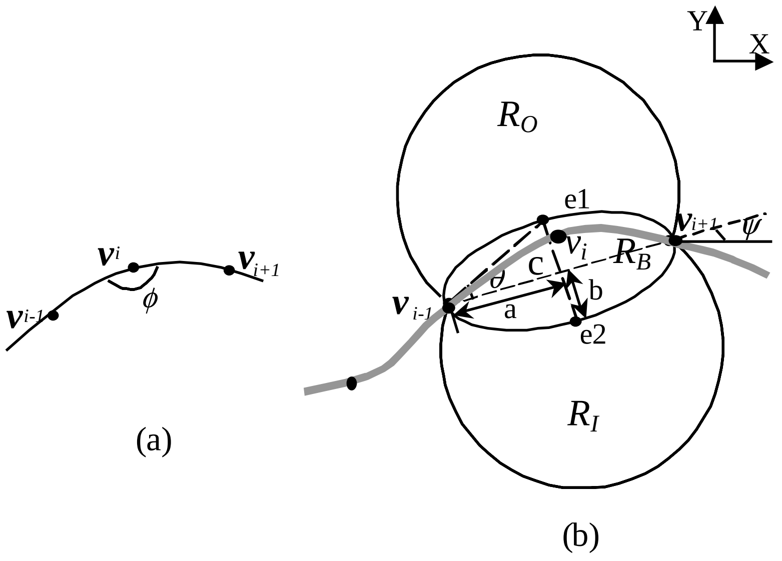
In this research, we propose a reformulated snake model, called robust snake model to confront the problems listed above. To decrease the sensitivity of contour initialization and the ambiguity from images, the new model combines the region and edge information into segmentation by employing a region model. Based on the framework of the region model, we reformulate both the internal and external energies of the traditional snake model, and the force function of level set models. As expected, its implementation shows it robust, fast and accurate.

In this section, we first introduce the construction of the region model and analyze its statistics, then we focus on how to employ it to reformulate the traditional snake model and level set model.

3.4.1 Region Model and its Statistics

The primary purpose of designing the region model is to estimate the region information around the snake contour, then use it to determine the direction of a normal force working on the snake model. This is very helpful to decrease the sensitivity of contour initialization.

The proposed region model is composed by three regions around the snake contour,



(a)

(b)

Fig 3-2: The construction of a region model

which are the outside region, the inside region and the boundary region. The construction of these three regions is based on the geometry of the current snake contour segment. To explain it more clearly, we use the snake contour in Fig. 3-2 as an example. The snake contour is sampled by a set of snaxels. To keep the contour smooth, the internal angle of each snaxel should be within a smoothness range $[\phi ,\pi]$, as shown in Fig. 3-2 (a). In the case of a closed snake which should behavior like a circle without the external energy, thus ϕ is defined as:

$$\phi = \frac{(N-2)\pi}{N} \tag{3-15}$$

Given the position of the previous snaxel $\mathbf{v}_{i-1}$, the ahead snaxel $\mathbf{v}_{i+1}$ and the sample number *N*, an ellipse shown in Fig. 3-2(b) can be constructed by calculating the following geometrical parameters:

$$\theta = \pi/N \tag{3-16}$$

$$\boldsymbol{c} = \mathbf{v}_{i-1} + (\mathbf{v}_{i+1} - \mathbf{v}_{i-1})/2 \tag{3-17}$$

$$a = |\boldsymbol{v}_{i-1} - \boldsymbol{c}| \tag{3-18}$$

$$b = a \bullet \tan(\theta) \tag{3-19}$$

$$\psi = \arctan\left(\frac{y_{i+1} - y_{i-1}}{x_{i+1} - x_{i-1}}\right) \tag{3-20}$$

where ***c*** is the center of the ellipse, *a* and *b* represent the length on main and minor axis, respectively. ψ denotes the direction of the main axis. We define the region enclosed by this ellipse as the boundary region (R_B), because most parts of a snake contour locate inside this region. To locate the other two regions, we use two feature points ***e***1 and ***e***2, the crossing points of the ellipse contour and its minor axis. The following equations can be used to compute ***e***1,***e***2:

$$M = \begin{bmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{bmatrix} \tag{3-21}$$

$$\boldsymbol{e}1 = M^{-1} (b \;\; 0)^T + \boldsymbol{c} \tag{3-22}$$

$$\boldsymbol{e}2 = M^{-1} (-b \;\; 0)^T + \boldsymbol{c} \tag{3-23}$$

where ***M*** is a rotation matrix. The outside region (R_O) here is defined as a region enclosed by a circle centering at ***e***1 with a radium $r = \sqrt{a^2 + b^2}$. Similarly, the inside region (R_I) is enclosed by the same-sized circle centering at ***e***2. Thus, if we trace along the snake contour in clock-wise direction, the outside regions are always on the left side of snake contours and the inside regions on the right side of snake contours.

In order to simplify the notation and to better introduce the proposed model, Some definitions are made:

H_S: the observed histogram of the region in the input image enclosed by the snake contour ***s***.

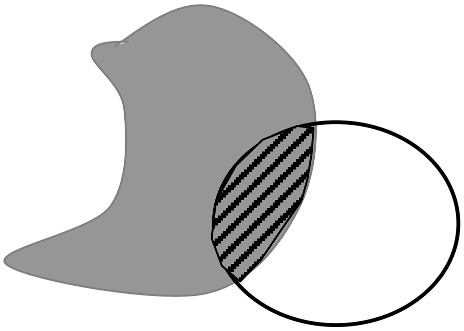


Figure 3-3: The definition of ROI

— : the snake contour

■ : the interested object

▨ : ROI

H_I: the histogram associated with the inside region R_I of a snake contour segment.

H_O: the histogram associated with the outside region R_O of a snake contour segment.

A key hypothesis made to perform segmentation is that we assume the intensity properties of Region Of Interest (ROI) can be represented using a Gaussian distribution. Generally speaking, there are many ways to define ROI according to different application purposes. In this study, we define ROI as:

$$ROI = \textit{Region enclosed by snake} \cap \textit{the interested object}$$

Figure 3-3 illustrates ROI for a given snake contour and its interested object. ROI is dynamic during the snake deforming. Ideally, it will become stable when the snake converges to the interested object boundary. Until that time, ROI is expected to be the region of the interest object. Due to its dynamic property, we need to recalculate ROI statistics after a certain number iterations to ensure the accurate estimation of ROI distribution.

Let $\mathbf{s}_i$ be a snake contour segment between $\mathbf{v}_{i-1}$ and $\mathbf{v}_{i+1}$. We estimate its inside and outside region probability by using their mean values μ according to Eq. 3-24:

$$P_I(s)\, or\, P_O(s) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\sigma_{ROI}} e^{\frac{-(\mu-\mu_{ROI})^2}{2\sigma_{ROI}^2}} & \dfrac{|\mu-\mu_{ROI}|}{\sigma_{ROI}} < t \\ 0 & \dfrac{|\mu-\mu_{ROI}|}{\sigma_{ROI}} \geq t \end{cases} \qquad (3-24)$$

Where the intensity mean μ_{ROI} and the standard deviation σ_{ROI} are computed from ROI. The constant t corresponds 90% confidence interval in a normal distribution. If the intensity mean of an insider or outside region falls within the pre-defined confidence interval of the statistical model of ROI, we set its probability by using the gaussian distribution function; otherwise we treat its probability as zero.

Let k_R be a measurement to evaluate the region property.

$$k_R(s) = \begin{cases} 1 & P_I(s) > 0 \quad and \quad P_O(s) > 0 \\ 0 & others \\ -1 & P_I(s) = 0 \quad and \quad P_O(s) = 0 \end{cases} \qquad (3-25)$$

To the observed snake segment $\boldsymbol{s}_i$, if its inside and outside region probabilities are both greater than zero, it is quite possible that this snake segment locates inside ROI. In such situation, a inflate force is needed to push the snake segment outside to enclose more homogeneous region. If its inside and outside region probabilities are both equal to zero, which means they don’t belong to ROI, a deflate force is used to pull the snake segment inside. Otherwise, we assume the snake segment is close to the boundary of ROI, and eliminate the region force.

Figure 3-4 illustrates the external energy working on the contour model with only an

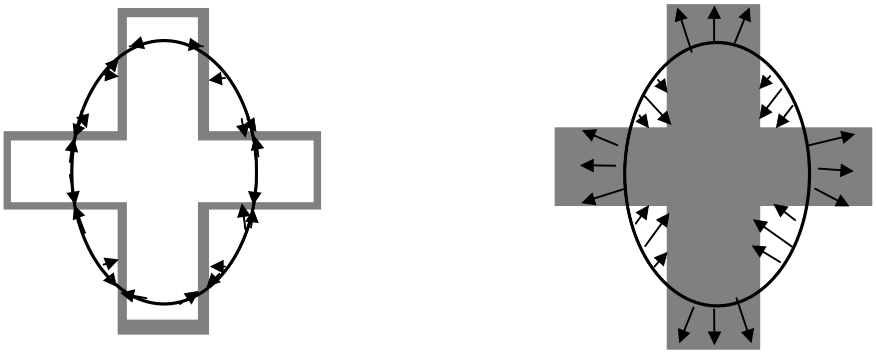

(a) (b)

Fig 3-4: The different external forces acting on the snake model

(a) Only edge-based forces. An edge map is estimated from the original image.

(b) Both edge-based and region-based forces.

edge-based force (Fig 3-4(a)) and with both region and edge force(Fig 3-4(b)). It is obvious that integrating both region and edge information into the external energy greatly enlarge the capture range of snake model and improve the snake performance.

3.4.2 Reformulate Snake Model with Region Model

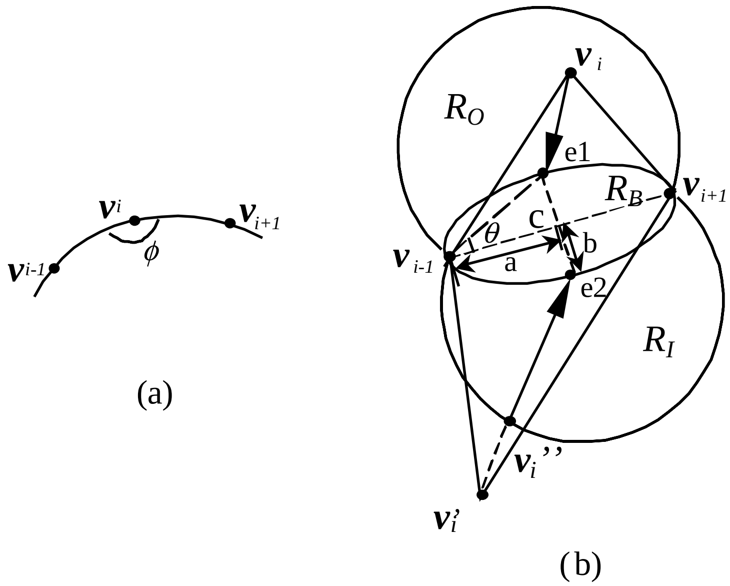
The proposed region model not only provides a reasonable way to generate the region force, but also helps to overcome the problems suffered by the internal energy of the traditional snake model. In this section, we focus on incorporating this region model to reformulate the traditional snake model.

Internal Energy

As mentioned earlier, the internal energy of the traditional snake model presents several deficiencies, such as the generation of contraction force, and not scale, translation and rotation invariant. Modifications of E_{int} can be divided into two classes. The first class focuses on removing the unwanted shrinking force by adding a new pressure force such as [Cohen 91, Sullivan 93]. However these methods add complexity to the snake's implementation and make its performance unpredictable, because of its difficult of determining a proper weight and correct direction of the pressure force. The second class promotes a new internal energy model based on the simple relationship between the local geometry of the contour and its internal parameters [Lai 94, Gunn 93, Li 99]. The new internal energy retains the conventional continuity and smoothness constraints, but replaces the shrinking force with an adaptive driving force. At the same time, the new model avoids the difficulty in adjusting proper weights of two derivative terms.

Based on the region model, we reformulate the internal energy of the snake model, which is somewhat similar to the second class of modification. As we know, the strength of the internal force of a snaxel is hard to predict due to the dependency of its contour shape. In contrast, the range of most external energy terms based on the edge information can be estimated. These facts result in the difficulty of choosing appropriate weights between the internal and external energy terms, which is the reason why even through all edge points have the same external energy, the snake will converge to some edge points but overrun others. To fix this problem, we design a new internal energy which generate "just enough" force to smooth the contour curve, overcome the external image forces and also enable to be normalized as well.

The new internal energy model determines its force based on three regions in the region model. According to the definition of the boundary region, any snaxel $\boldsymbol{v}_i$ inside



(a)

(b)

Fig.3-5: The internal energy model

this ellipse region would satisfy with the smooth constraint. Thus, we treat its internal energy as zero.

For a snaxel outside the boundary region, but locating in either the outside or the inside region, its internal force should pull it into the boundary region and also ensure it evenly spaced between its previous and ahead points. To compute this force E_{int}, we use two feature points ***e1*** and ***e2***. The advantages of choosing these two points are that, firstly, they locate at the edge of the boundary region and satisfy the minimal smooth angle constraint. Secondly, they have the same distances between its previous and ahead snaxels so that the tension forces from these two points are equal. Thus they are the best positions to maintain both smoothness and tension constraints. The internal force here can be defined as a force pulling a snaxel to its nearer feature point, such as $\boldsymbol{v}_i$ in Fig 3-5 (b).

For the snaxel outside of all defined regions, the internal force might be very strong because of high curvature. To avoid the internal force overact, we constraint its force magnitude equal to *th*, but still keep its original direction. ***v'***i in Fig 3-5 (b) illustrates such situation. Accordingly, The internal force can be defined as:

$$E_{\text{int}}(v_i) = \begin{cases} 0 & v_i \in \quad R_B \\ \dfrac{1}{th}\min\limits_{j=1,2}(v_i - e_j) & v_i \in \quad R_I \quad or \quad R_O \\ \dfrac{1}{|\min\limits_{j=1,2}(v_i - e_j)|}\min\limits_{j=1,2}(v_i - e_j) & v_i \in \quad others \end{cases} \qquad (3-26)$$

Advantages of the new internal energy over the traditional snake include 1) It provides no prejudice towards expansion or contraction. So it can be used along with either an inflating or a deflating force. 2) It is proven to be scale rotation and translation invariant. Thus, it suits to represent both concave and convex curve shape of contours.

PROOF: The reformulated internal energy is scale, rotation and translation invariant.

Suppose a translation, rotation and scale transformation can be defined as:

$$u = s\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} v + \begin{pmatrix} x_{offset} \\ y_{offset} \end{pmatrix} = Av + B \qquad (3-27)$$

From (3-26), the internal energy is,

$$E_{\text{int}}(u_i) = \frac{\min\limits_{j=1,2}(u_i - e_j)}{\sqrt{a(u_i)^2 + b(u_i)^2}} \qquad (3\text{-}28)$$

For simplicity, we only use ***e***1 for proof, e2 is similar.

$$E_{\text{int}}(u_i) = \frac{u_i - e1}{\sqrt{a(u_i)^2 + b(u_i)^2}} = \frac{u_i - M^{-1}(1 \quad 0)^T b(u_i) - c(u_i)}{\sqrt{a(u_i)^2 + b(u_i)^2}} \qquad (3\text{-}29)$$

For each item in (3-29)

$$c(u_i) = u_{i-1} + \frac{(u_{i+1} - u_{i-1})}{2} = Av_{i-1} + B + \frac{A(v_{i+1} - v_{i-1})}{2}$$

$$= A\left[v_{i-1} + \frac{(v_{i+1} - v_{i-1})}{2}\right] + B$$

$$Thus: \quad c(u_i) = Ac(v_i) + B$$

$a(u_i) = |u_{i-1} - c(u_i)| = |Av_{i-1} + B - [Ac(v_i) + B]| = A|v_{i-1} - c(v_i)|$

Thus: $a(u_i) = A \bullet a(v_i)$

$b(u_i) = a(u_i) \bullet \tan(\theta) = A \bullet a(v_i) \bullet \tan(\theta)$

Thus: $b(u_i) = A \bullet b(v_i)$

Since $M(\mathbf{u}_i) = M(\mathbf{v}_i)$.

$$E_{\text{int}}(u_i) = \frac{(Av_i + B - M^{-1}(1 \quad 0)^T Ab(v_i) - Ac(v_i) - B)}{\sqrt{[Aa(v_i)]^2 + [Ab(v_i)]^2}}$$

$$= \frac{A\left(v_i - M^{-1}(1 \quad 0)^T b(v_i) - c(v_i)\right)}{A\sqrt{[a(v_i)]^2 + [b(v_i)]^2}}$$

$$= E_{\text{int}}(v_i)$$

External Energy

The proposed new external energy composes two forces: a region pressure force and an edge force. If a snaxel's neighbor fits statistical model of ROI, a region pressure force normal to the snake curve will push it outside. Conversely, if it lies outside of the model of ROI, the region pressure force is reversed to pull it inside. Only when the snake is close to the object boundary, the edge force will work on the snake. Thus, the external energy can be given by:

$$E_{ext} = (1-k_R)\, E_{edge} + k_R\, E_{reg} \qquad (3\text{-}30)$$

Another important fact is that, after modification, both the internal and external energy can be normalized. For the sake of simplification, we just assign same weights to these two energy terms.

3.4.3 Reformulate Level Set Models with Region Model

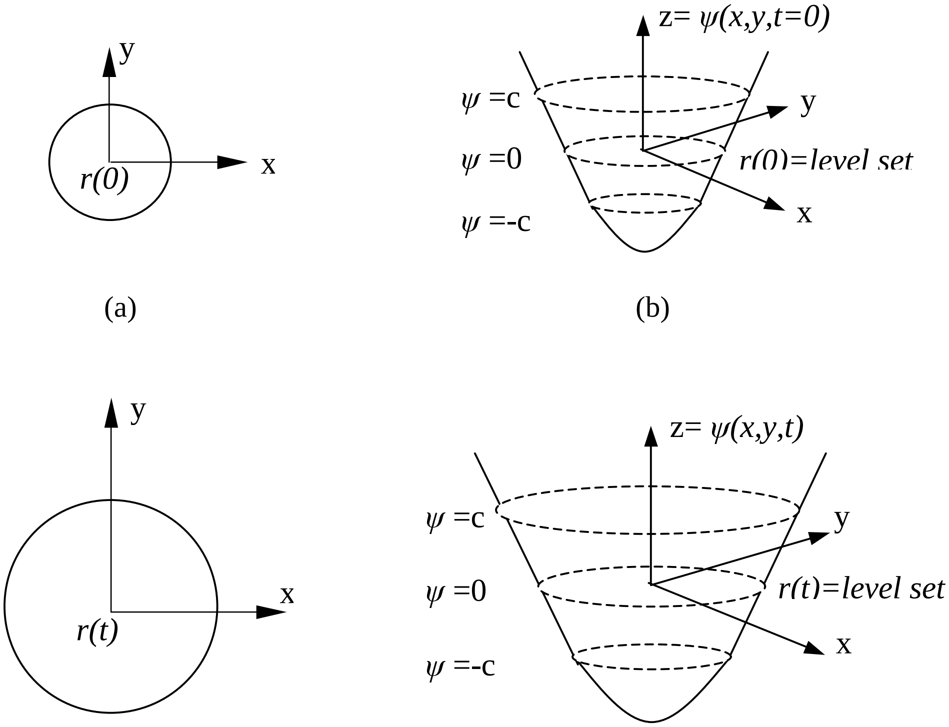
The use of the level set methods, introduced by Osher and Sethian [Osher 95] provides a very elegant tool to propagate curve (or front) along its gradient field with constant speed or curvature-dependent speed.

The central idea in the level set approach is to represent the front $\boldsymbol{r}(t)$ as the zero-level set $\{\psi=0\}$. Thus, given a moving closed hypersurface $\boldsymbol{r}(t)$, that is, $\boldsymbol{r}(t=0):[0, \infty] \rightarrow R^N$, they wish to produce an Eulerian formulation for the motion of the hypersurface propagating along its normal direction with speed $\boldsymbol{F}$, where $\boldsymbol{F}$ can be a function of various arguments, including the curvature, normal direction, etc. The main idea is to embed this propagating interface as the zero level set of a higher dimensional function ψ. Let $x(s,t)$ be the position vector which parameterizes $r(t)$ by s, $0 \leqslant s \leqslant S$

$$\psi(x(s,t),t=0) = \pm d \qquad (3\text{-}31)$$

where d is the distance from $\boldsymbol{x}$ to the initial hypersurface $\boldsymbol{r}(t=0)$, and the plus(minus) sign is chosen if the point $\boldsymbol{x}$ is outside(inside) of $\boldsymbol{r}(t=0)$.

As illustration, consider the example of an expanding circle. Suppose the initial front $\boldsymbol{r}$ at t=0 is a circle in the xy-plane (Fig 3-6(a)). We imagine that the circle is the level set $\{\psi=0\}$ of an initial surface $z=\psi(x,y,t=0)$ in R^3 (see Fig 3-6 (b)). We can then match the one-parameter family of moving curves $\boldsymbol{r}(t)$ with a one-parameter family of moving surfaces in such a way that the level set $\{\psi=0\}$ always yields the moving front (see Fig3-



(a) (b)

(c) (d)

Fig 3-6: Level Set formulation of motion

6(c) and Fig. 3-6 (d)).

Based on this idea, Osher and Sethian define the evolution equation for ***ψ**(x,t)* which contains the embedded motion of ***r**(t)* as the level set *{**ψ**=0}*:

$$\Psi_t+F|\nabla\psi|=0 \qquad (3\text{-}32)$$

To implement the level set formulation for shape recovery, the front can be represented by the boundary of an evolving shape. The speed function $\boldsymbol{F}$ is spitted into two components: $\boldsymbol{F}=\boldsymbol{F}_A+\boldsymbol{F}_G$. The term $\boldsymbol{F}_A$ referred as the advection term is independent of the moving front geometry. The front uniformly expands or contracts with speed $\boldsymbol{F}_A$ depending on its sign and is analogous to the inflation force defined in balloon snake. The term $\boldsymbol{F}_G$ depends on the geometry of the front, such as its local curvature. This diffusion term smoothes out the high curvature regions of the front and has the same regularizing effect on the front as internal deformation energy term in thin-plate-membrane spline in Kass snake. Thus Equation 3-32 can be rewritten as:

$$\Psi_t + k_I (F_A + F_G) |\nabla \psi| = 0 \qquad (3\text{-}33)$$

$$k_I(x,y) = 1 / |1 + |\nabla G_\sigma * I(x,y)| \qquad (3\text{-}34)$$

Where k_I is an image related coefficient which has value closer to zero in regions of high image gradient and values closer to unity in regions with relatively constant intensity. The final shape can be the configuration when all the points on the front come to a stop, thereby, bringing computation to an end.

According to the equation, the speed function in level set method can be positive in some place and negative in others, so that the front can move forward in some place and backward in others. Although it gives the method tremendous generality, it significantly slows down the processing speed. Instead of searching for the correct direction of the speed function, we derive it by using the region information. As mentioned in section 3.2, the region model provide a useful tool to predict if the front is inside or outside the object. With the reference of its statistical information, the reformulated equation is given by

$$\Psi_t + k_I (k_R F_A + F_G) |\nabla \psi| = 0 \qquad (3\text{-}35)$$

Thus, the reformulate speed function combines both edge and region information. The region information shrinks or expands the front curve to the direction which enclosed ROI as much as possible. The edge information attracts the curve towards the boundaries of ROI.

3.5 Performance Assessment

The proposed deformable model has been applied on both synthetic and real gray-level images.

3.5.1 Experimental Results on Simulated Images

To demonstrate the performance, our method is compared with a traditional Kass snake [Kass 87], the balloon snake[Cohen 91], and GVF snake[Xu 98].

This experiment tests on a synthetic image shown in Fig.3-7(a) which contains two objects, a white and a gray object with gaussian noise. We firstly define a circle shown in Fig.3-7(a) as the initial contour and its center points in the white object which is the interested object. Starting from the same initial contour, the final snake contours detected by a tranditional snake, the balloon snake and GVF snake are shown in Fig.3-7(b)(c)(d), respectively. We notice that for this initialization, the convention snake (Fig.3-7(b)) only works around the initial snake. The balloon snake with outward pressure forces causes the snake to expand into some boundaries of objects. But it can not detect the boundaries inside the initial contour as shown in Fig.3-7(c). The GVF snake result is shown in Fig.3-7(d) .It has converged to some parts of object boundaries. However it is still affected by the initial contour. Since the snaxels only move to the nearest boundaries, for those boundaries which are not the nearest to any snaxels, no snaxle can converge to them. Thus, GVF snake fails to find all boundaries of object. Another major problem is that all above snakes only converge to nearby boundaries no matter whether they belong to the interested object or not. The reformulated snake models fix all above problems and

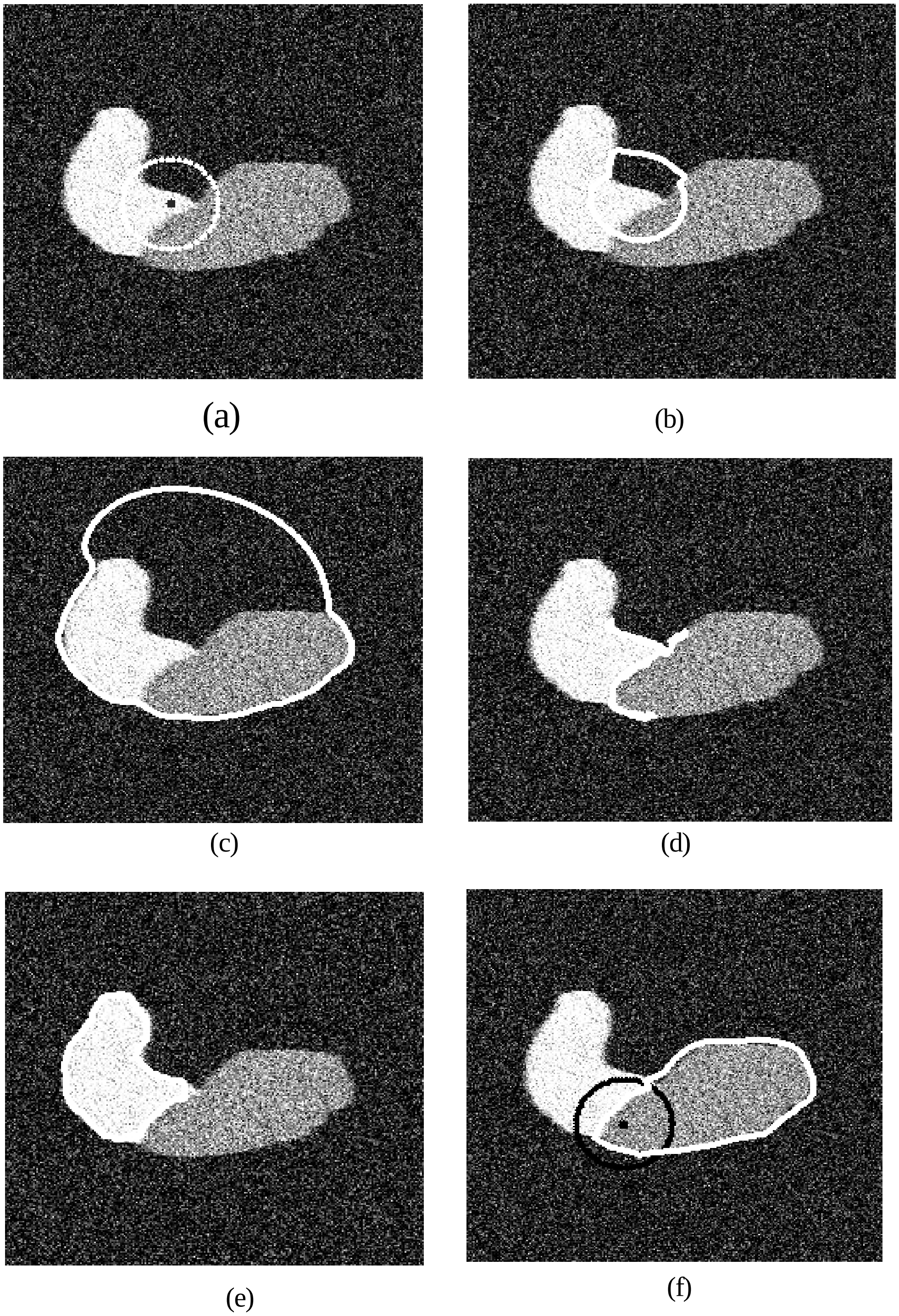
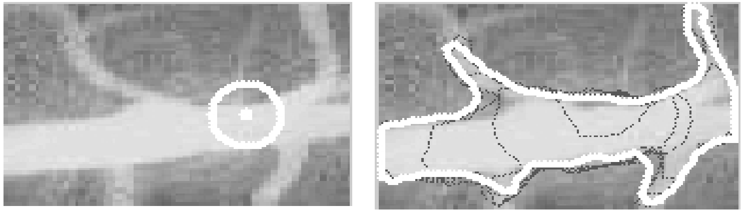

(a) (b) (c) (d) (e) (f)

Fig.3-7: The comparisons of different methods working on a synthetic image clearly obtain the best result (Fig. 3-7(e)). Fig.3-7(f) shows the result (white) used to detect the gray object from the initial contour (black). As expected, the proposed snake models appear insensitive to initialization since it is clear when to apply inward or outward pressure forces. For the reformulated level set method, the experimental results are exact same as Fig 3-7 (e) and (f).

3.5.2 Experimental Results on Real Images

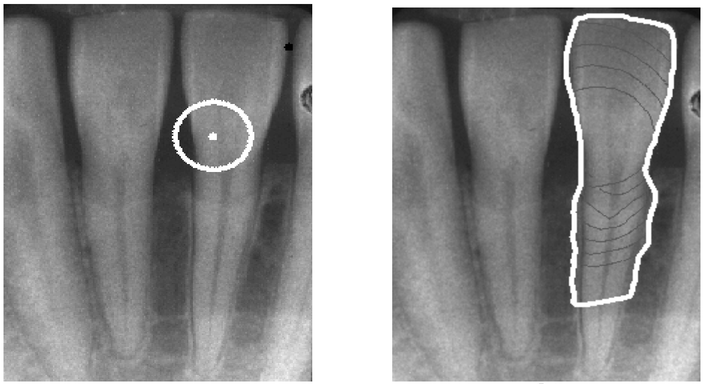
In this section, we demonstrate the performance of the proposed snake models working on the real images. The selected images includes a DSA image (Fig 3-8), an X-ray image (Fig 3-9), chest radiographs (Fig 3-10) and radiographs of different body parts (Fig 3-11,12). In chest radiographs, we focus on the detection of lung regions. It can be observed that the lung contour can be accurately located by using our robust snake method. As mentioned earlier, this research aims to automatically analyze and classify radiographs. The proposed snake models, as an important part of the system, is to extract the contour of the interested body object from radiographs. With the help of the contour information, the related parameters, such as position, size and orientation, can be estimated and used for model modification. Experimental results from Fig. 3-11,12 and 13 demonstrate the proposed snake model satisfy with our system requirements and most of its results are similar to the demonstrated results in Fig 3-11,12 and 13.

.



(a) (b)

fig. 3-8: The experimental result of a DSA image



(a) (b)

fig 3-9: The experimental result of a X-ray teeth image

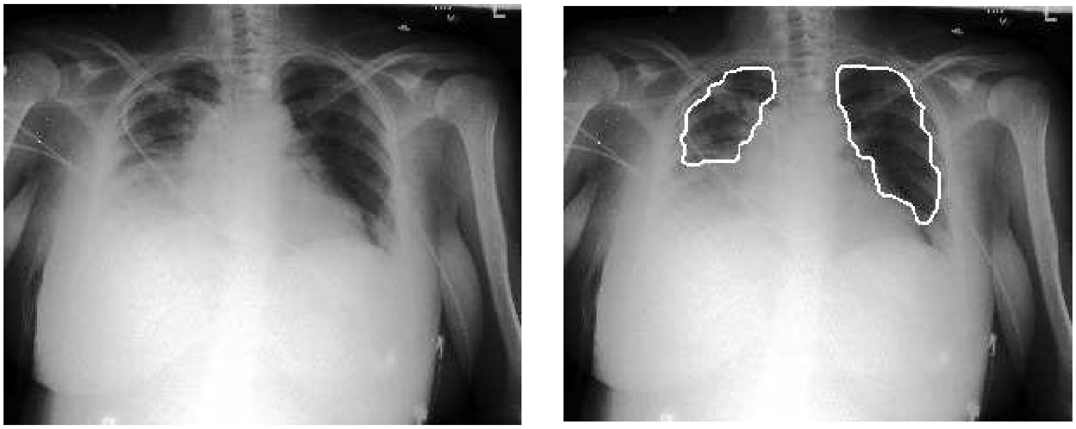


Fig 3-10: The lung region detection of a chest radiograph

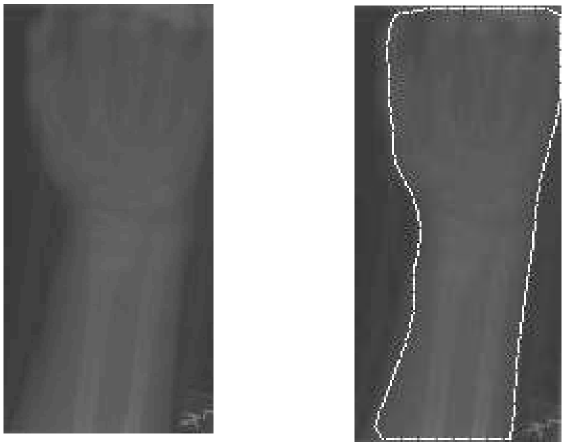


Fig 3-11: The contour detection of a wrist radiograph

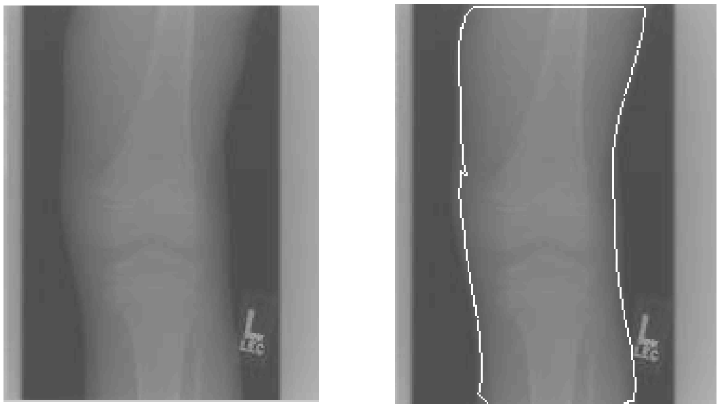


Fig 3-12: The contour detection of a knee radiograph

Fig 3-13: The contour detection of an ankle radiograph

3.6 Summary

In this chapter, we discuss the significant properties of the snake model and the problems of implementing the model. With the motivation of improving the behavior of the snake model, we proposed a region model to combine both region and edge information together in the segmentation, which directly overcome the problems associated with the initialization, then we reformulate the traditional snake model and level set models with it. As expected, the performance of the reformulated models is greatly improved and experimental results are encouraging.

Chapter 4:

Architecture of Content-based Medical Image Classification (CMIC) System

4.1 Introduction

The rapid convergence of medical imaging technology into digital acquisition and storage has made digital Picture Archiving and Communication System (PACS) possible. These systems, important to hospitals, are growing at a significant rate. The frequency with which medical imaging is used in patient diagnosis has resulted in substantially large image collections. Even modest size hospitals may create, and need to store more than a million individual images annually. However, medical image classification and management systems, to date, have been primarily based on a few textual or numerical fields residing in standard relational database structures which is not sufficient to represent the information contained in images. There is reason to believe that the performance of medical image management system might be greatly improved if image databases can be organized and retrieved based on image content. To organize by image content implies the detection of anatomical structures in images as well as the understanding of their geometry.

However, Image data is more structurally complex than other kinds of data. It is hard to let the computer automatically and efficiently analyze contents in images, since the way with which human beings capture the image content, group image features into meaningful objects and attach semantic descriptions to images through model matching has not been fully understood to automate the analysis procedure. Segmenting an image into regions corresponding to individual objects, extracting features from the image that capture the perceptual and semantic meanings and matching the image with the proposed model based on the extracted features also make the analysis problem more challenging.

Various systems have been proposed in the recent literatures for content-based image classification and retrieval, such as QBIC [Niblack 94], Photobook [Pentland 96], Virage [Bach 96], Visualseek [Smith 96], Netra [Ma 97], and MAR [Huang 96]. These systems follow the same paradigm which treats an image as a whole entity and represents it via a set of low-level feature attributes, such as color, texture, shape and layout. As a result, all these feature attributes together form a feature vector for each image. The image classification is based on clustering these low-level visual feature vectors. Such clustering-based classification schemes are usually time-consuming and of limited practical use since little of the image object semantics is explicitly modeled.

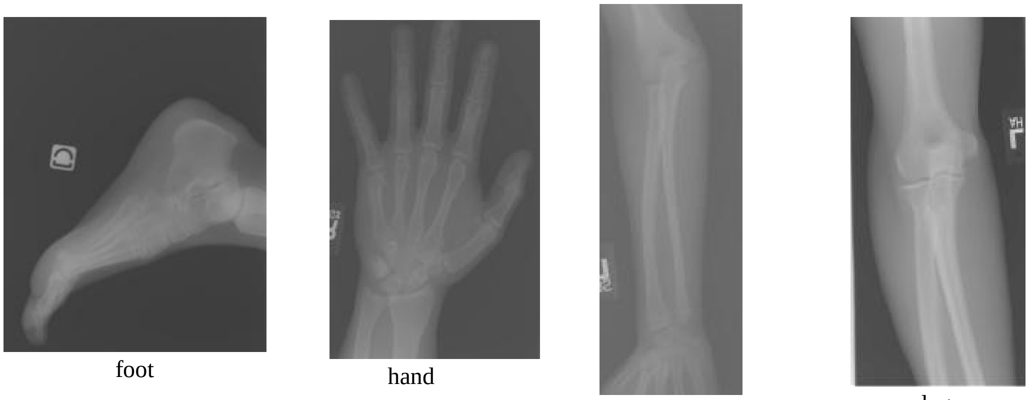
We argue that understanding the entire content of an image may not be sufficient with such simple low-level feature extraction and matching algorithms in computer vision, since the fact that images have hierarchical structure is ignored. An experimental study conducted by Nishiyama et al [Nishiyama 94] points out that human beings have two patterns in their visual memory when they view images. The first pattern is to roughly view the whole image and focus on the object outlines that are visually or semantically interesting. The second pattern is to concentrate on detailed attributes of specific objects within the image. Thus, to design an efficient and effective content-based medical image classification system, we need to appropriately combine such hieratical structure with feature extraction and match algorithms.

4.1.1 Characteristic of Medical images

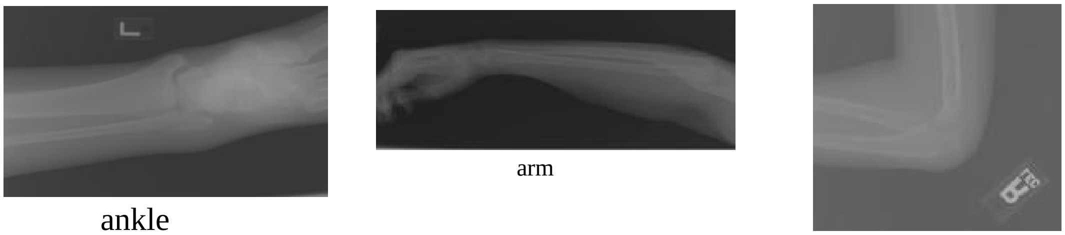
Compared to the general image databases, the medical image database has the following characteristics:

1) *The inherent ambiguity of radiographs*: A radiograph is a transmission image of objects by means of X-rays. Due to the limitations of image acquisition parameters, the image contrast in computer radiographs is low. This problem might be further complicated if an image consists of several regions with different X-ray attenuation characteristics. In addition, all the procedures in image acquisition, such as the instrumentation precision, the sampling resolution, visual perspective and lighting may also introduce noise in image objects and distort their morphological and geometric characteristics.

2) *The complicate anatomy structure:* As shown in Fig 4-1, different body parts in radiographs demonstrate a high degree of variation. Even to the same structure, it will also demonstrate different characteristics because of its location and view angle. Besides of these, the presence of overlapping or adjacent bone structures and the large variability



foot hand forearm leg



ankle arm knee

pelvis skull abdomen

shoulder knee chest

Fig 4-1: Sample images from medical image databases of bone structures, according to age, pose and anatomy of individual may also influence the accuracy of detection.

3) *The priori knowledge:* Unlike general image databases, we have the priori knowledge of contents in the medical image database, which implies what kinds of object are expected in the processed images.

These three characteristics have a tremendous influence on the design of content-based medical image classification system. The ambiguous nature of medical images implies that the system should have capability to handle uncertainty. The fact that anatomical structures have a high degree of variation imposes several requirements on the knowledge model.

4.1.2 Problem Definition

The image classification problem of interest here can be defined as follows: Given an image, the system must derive appropriate descriptions of the image contents and classify it into one image category of human body parts based on the detected contents. In this research, we assume that each input image belongs to one of pre-defined image categories. However, in real world images, because of the ambiguity of images and the similarity among anatomical knowledge models, it is quite possible that the input image can be placed into one or more image categories. To reduce such ambiguity, we carefully study the processed results and reject those unreasonable classifications with the help of numerous correct classification results. Thus, the outputs of the system are the descriptions of image contents and one or more image categories which have high confidence levels assigned by the system.

The rest of this chapter is organized as follows: Section 4.2 briefly overviews the architecture of content-based medical image classification system. To address the system systematically, a detail description of coarse shape classification is presented in Section 4.3. Section 4.4 discusses how to implement a detail content identification on the processed images. Then, the performance of the system is evaluated in section 4.5. Section 4.6 shows some experimental results, followed by the conclusion in section 4.7.

4.2 Framework of the System

To solve the problem, the proposed object-oriented knowledge model provides a powerful means for medical image understanding and classification. The key idea here is that: content features of medical images must hierarchically match to the related knowledge model. Based on the characteristics of the model, the hypotheses used for the classification in the system can be described as:

1) For each image category, the database system should maintain a knowledge model class which encapsulates all features relevant to this category and methods to detect these features from the processed images.
2) During classification, one knowledge model class can be assigned to an image only when the image features detected by the model class methods satisfy the pre-defined model “deformability” limits.

This strategy is likely to yield far better results than using a general strategy which classifies an image only by a set of low-level image features, since it exploits knowledge not only regarding the characteristics of one image category but also the ways to detect

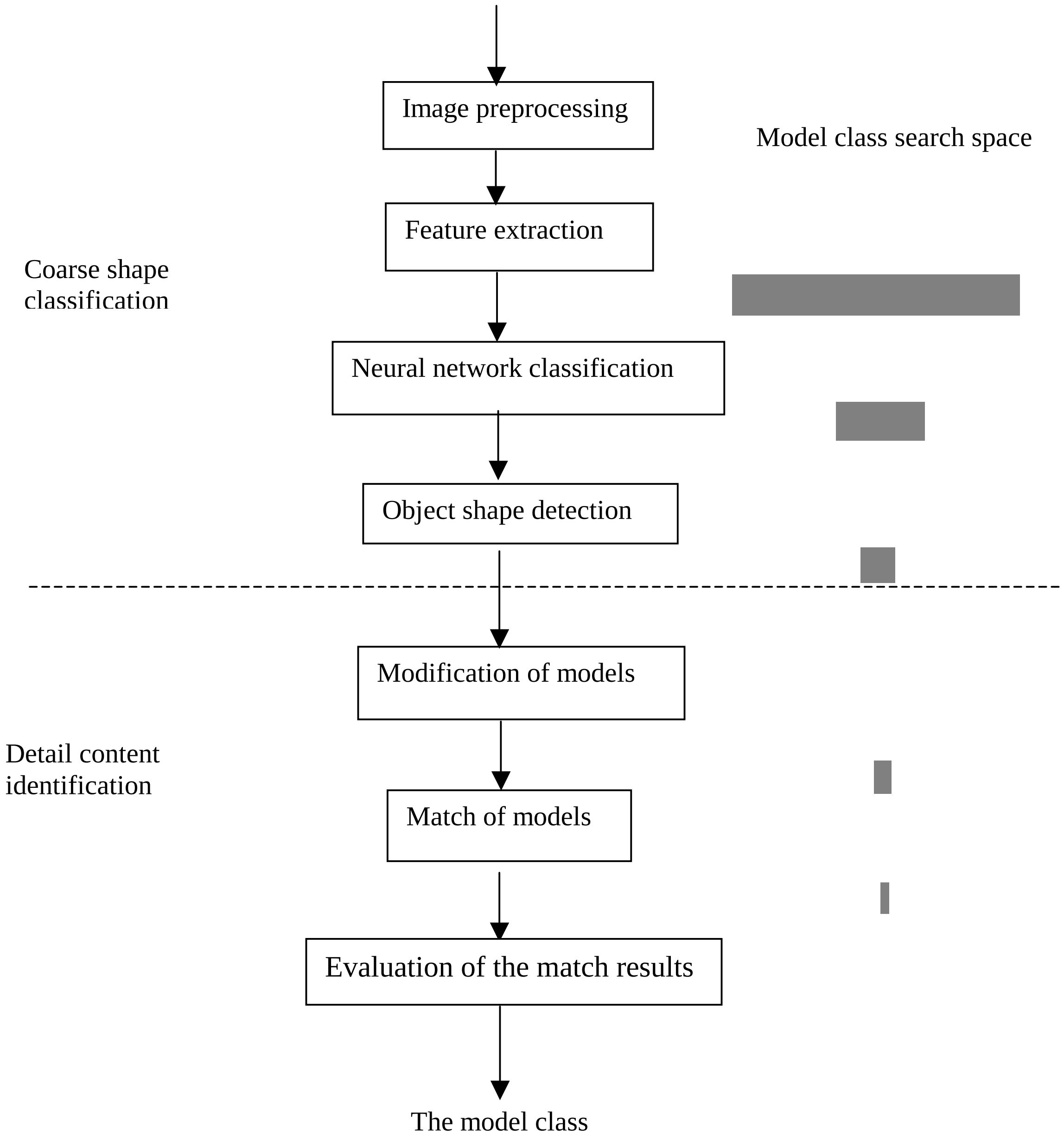


Fig 4-2: Architecture of the classification system

and recognize them.

What we propose here is a two-stage hierarchical image classification system. Fig 4-2 shows the architecture of the system prototype. The first stage focuses on the coarse pattern of the model class and ignores all finer details of individual parts. So only description attributes of the main object in the model class are used for classification, such as shape. To accomplish it, this stage performs the following functions: image preprocessing, feature extraction, neural network classification and object shape detection and classification. After the coarse shape classification, the system classifies the input image into one of three clusters mainly based on the shape characteristics of the interested object. The importance of this stage is to reduce the number of models needed for detail matching and speed up the processing. Once the coarse shape classification is done, a small set of plausible model class candidates are then presented for a detailed matching in the second stage. In this stage, the local features used to identify salient component objects in images are employed and incorporated with their spatial relation to discard those spurious match candidates. The output match results imply a set of model which might be existed in the processed images. To further confirm them, an evaluation strategy is used to reject those unreasonable results and inference the most possible match models.

4.3 Coarse Shape Classification

As we mentioned earlier, the classification in this stage focuses on the description attributes of the main object. Based on the characteristics of our medical image database, the only attribute that can be applied for the discrimination is the shape outline of the main object. However, incorporating scale, rotation and translation invariance in shape matching generally increases the computational requirements. So, the aim of this stage is to design a fast and robust scheme for shape classification.

4.3.1 Image Preprocessing

To speed up the processing, one simple way is to reduce the size of images by using the pyramid data structure. In this study, we set the original image as the lowest level of the pyramid. Each higher level pyramid is derived by reducing one-half resolution of the lower level image. If in a certain level, the size of the pyramid image is less than 50 pixels, the pyramid creation is stopped. The obtained image in this level will be used as the input image for coarse shape extraction.

Since what we only concern about is the shape outline of the main object in images. So to isolate the shape from its background, we employ the maximum entropy thresholding technique to automatically compute a threshold from the histogram of the processed image.

The maximum entropy sum method proposed by Kapur et al [Kapur 85] is based on the maximization of the information measure between two classes (object and background). Consider a discrete image with $N*M$ pixels defining the object and background. Let $p_i = h(i) = n_i / NM$ be an estimate of the probability of its gray-level value, where n_i represents the number of pixels with gray value i, The *a priori* entropy of the entire image is given by

$$H_T = -\sum_{i=0}^{255} p_i \ln p_i \qquad (4-1)$$

If we assume two classes of pixels, *C1* and *C2*. *C1* denotes the class of object pixels and *C2* denotes the class of background pixels. In the image, the *a priori* entropies are defined as

$$H_{C1}(t) = -\sum_{i=0}^{t} \frac{p_i}{p(C1)} \ln \frac{p_i}{p(C1)} \qquad (4-2)$$

$$H_{C2}(t) = -\sum_{i=t+1}^{255} \frac{p_i}{p(C2)} \ln \frac{p_i}{p(C2)} \qquad (4-3)$$

where

$$p(C1) = \sum_{i=0}^{t} p_i \qquad (4-4)$$

$$P(C2) = \sum_{i=t+1}^{255} p_i \qquad (4-5)$$

$$p(C1) + p(C2) = 1 \qquad (4-6)$$

The information between the two classes, *I(C1,C2)* is defined as

$$I(C1,C2) = H_{C1}(t) + H_{C2}(t) \qquad (4-7)$$

It can also be rewritten as a function, $\psi(t)$, of the threshold *t*. The above equation can be simplified as

$$\begin{aligned} \varphi(t) &= \frac{-\sum_{i=0}^{t} p_i[\ln p_i - \ln p(C1)]}{p(C1)} + \frac{-\sum_{i=t+1}^{255} p_i[\ln p_i - \ln p(C2)]}{p(C2)} \\ &= \ln p(C1) + \ln p(C2) + \frac{-\sum_{t=0}^{t} p_i \ln p_i}{p(C1)} + \frac{-\sum_{t=t+1}^{255} p_i \ln p_i}{1 - p(C1)} \\ &= \ln[p(C1)p(C2)] + \frac{H(t)}{p(C1)} + \frac{H_T - H(t)}{1 - p(C1)} \qquad (4-8) \end{aligned}$$

Where *H(t)* is defined as

$$H(t) = -\sum_{i=0}^{t} p_i \ln p_i \qquad (4-9)$$

The method due to Kapur et al. chooses the threshold to be the value *t**, at which the function $\psi(t)$ is maximum.

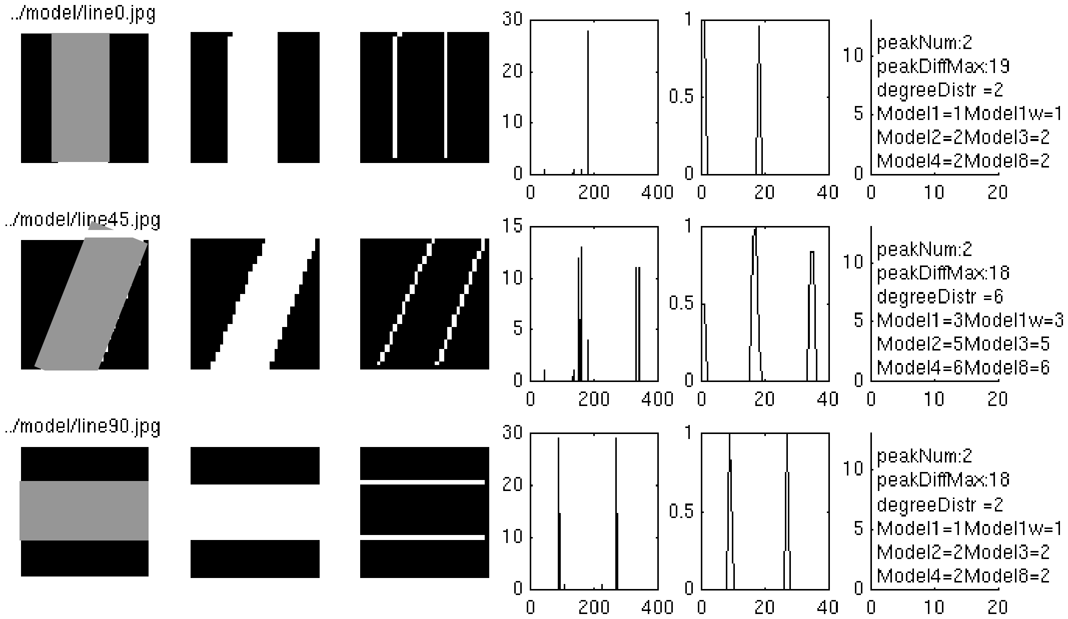
The binary image contains speckle and has some misclassified regions. This ‘noise’ can be easily removed using standard morphological operations. In this study, we use binary erosion first followed by dilation. After preprocessing, the white region in the binary image approximates the shape of the main object.

4.3.2 Feature extraction

A concise and quantitative description of the object shape is a challenging problem, especially describing a shape by a set of scale, rotation and translation invariant features. To achieve it, we extract a set of shape features from the histogram of the edge directions. The edge image is generated by using the sobel edge operator on binary images, and its corresponding edge directions are then quantified into 36 bins of 10^o each. To explain it more clearly, some synthetic image examples are shown in Fig 4-3.

There are many advantages and limitations in representing a shape with its edge directions. In the following parts of this section, we want to analyze the properties of edge directions and try to take use of its advantages and overcome its limitations to exact a set of scale, rotation and translation invariant features for classification.

- The histogram of edge directions is invariant to translations in an image. Thus the position of the main objects in the image has no effect on its edge directions. This property can ensure that any features extracted from the histogram of edge directions are translation invariant.
- The use of edge directions is inherently not scale invariant. Two images identical in every aspect except their size will yield different number of edge points and hence result in two totally different histograms. In order to achieve invariance to



(a) (b) (c) (d) (e) (f)

Fig 4.3: Illustration of the histogram of edge directions

(a): the original images
(b): the binary images of the original images
(c): the edge images of the original images
(d): the histogram of edge directions
(e): the quantified histogram of edge directions
(f): the nine scale, rotation and translation invariant features

scale, we normalize the histogram as follows:

$$H(i) = H(i)/n_e \qquad i \in [0,\ldots,35]$$

Where $H(i)$ is the count in bin i of the edge direction histogram, n_e is the total number of edge points detected in the image.

- A histogram of the edge directions is not invariant to rotation either. As shown in Fig.4-3, these three images contain an object with the same shape but different orientations. So they have different histograms of edge directions. However, if we only consider the shape feature of these histograms, they are quite similar in that all of them have two peaks with 180^{o} distance. If we can design a feature which only captures such information, the histogram of edge directions can be made invariant to rotation. To achieve this, we generate a set of edge direction histograms from the shape models which correspond to some typical shapes extracted from our medical image databases, and then use them as histogram templates. The match between the image object and typical model shapes is performed by first shifting the histogram templates, and then match them with the edge direction histogram of the processed image. The shift of histogram templates during matching can be seen as the rotation of their corresponding shape models.

- The matching results depend on the bin size. If choosing a larger bin size, for example 20^{o} each bin, the matching speed is improved. However, this also reduces the match accuracy because the image object may have an arbitrary rotation in the image. A rotation not only causes a shift in the histogram bins but also affects the bin membership. On the other hand, use of a smaller bin size can improve the accuracy of system, but sacrifices the matching speed. According the accuracy requirements of the system and a number of experimental results, we found that using 36 bins to represent edge directions can satisfy our requirements.

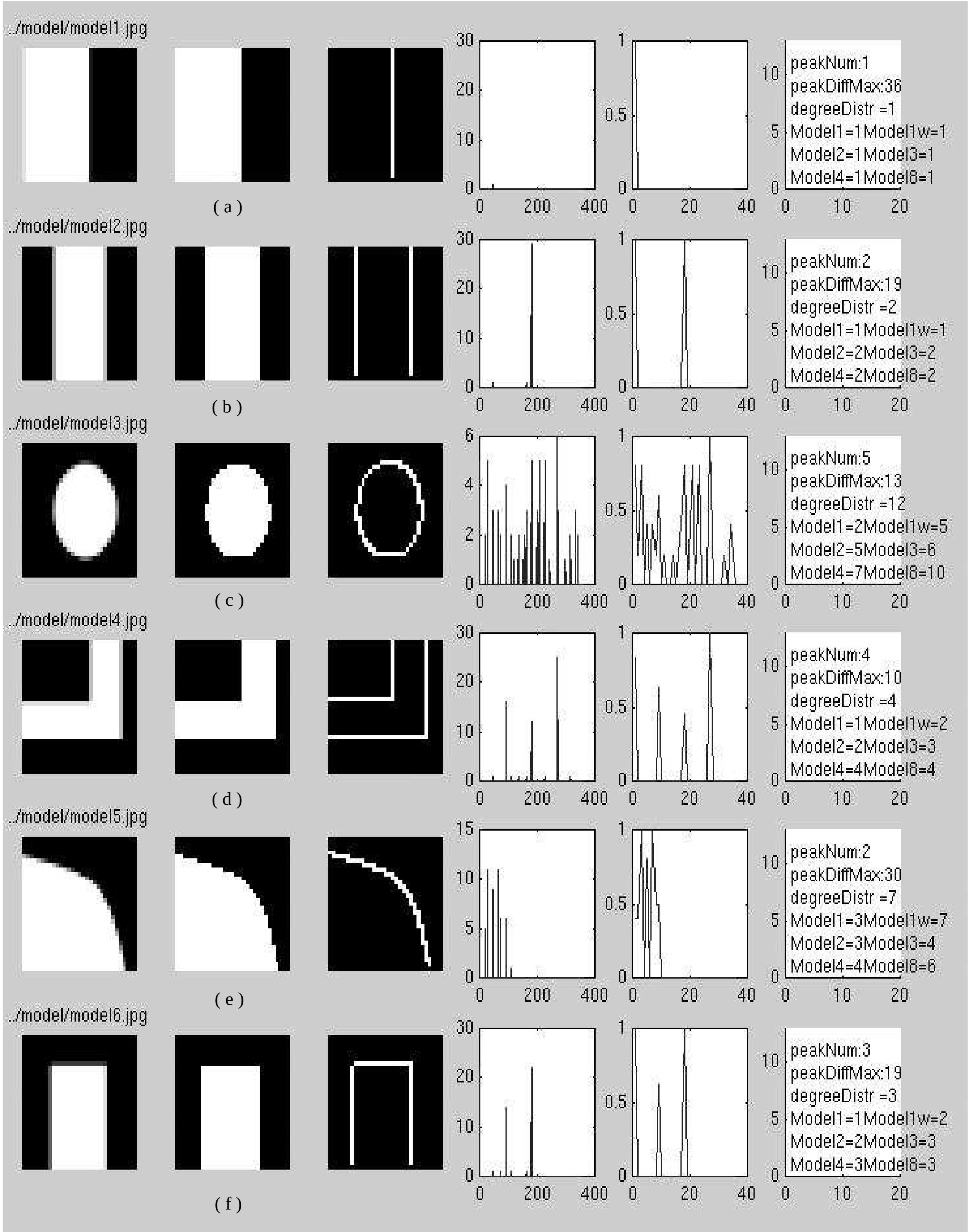
Selecting a good feature set to represent the image shape is a fundamental problem in image classification. Through carefully analyzing the characteristics of our medical image database and numerous experiments, we define three coarse shape classes based on the characteristics of edge direction histograms: the one-peak class, the two peaks class and the more peaks class. We evaluated the following nine features to discriminate these three classes:

- *The peak number: the number of peaks in the histogram of edge directions*
- *The minimal peak distance: the minimal distance between two nearest peaks*
- *The occupied degree number: the total number of degree contained in the histogram of edge directions. This feature is helpful to describe the degree distribution of different shape.*
- *The maximal match value of Model1: the maximal convolution value between model1 and the histogram of edge directions. Model1 shown in Fig 4-4(a) represents the shape with only one straight edge.*
- *The maximal match value of Model1w: the maximal convolution value between model1w and the histogram. Model1w shown in Fig 4-4(e) represents the shape with a curve edge.*
- *The maximal match value of Model2: the maximal convolution value between mode2 and the histogram. Model2 shown in Fig 4-4(b) represents the shape with two straight edges.*
- *The maximal match value of Model3: the maximal convolution value between model3 and the histogram. Model3 shown in Fig 4-4(f) represents the shape with three edges.*
- *The maximal match value of Model4: the maximal convolution value between model4 and the histogram. Model4 shown in Fig 4-4(d) represents the shape with four main edges.*
- *The maximal match value of Model8: the maximal convolution value between model8 and the histogram. Model8 shown in Fig 4-4(c) represents the shape like circles or ellipses.*

These six histogram templates include the main coarse shape features in our medical image database, and provide a good way to evaluate the similarity between the main object shape features in the model class and image shape features. Fig 4-5 shows six typical shape images from our medical image database and their corresponding shape features, which are similar with the six histogram templates in Fig 4-4, respectively.

4.3.3 Neural networks training

Neural networks have been widely used in pattern recognition. In this section, we



(a) (b) (c) (d) (e) (f)

Fig 4-4: The histogram template and their corresponding images

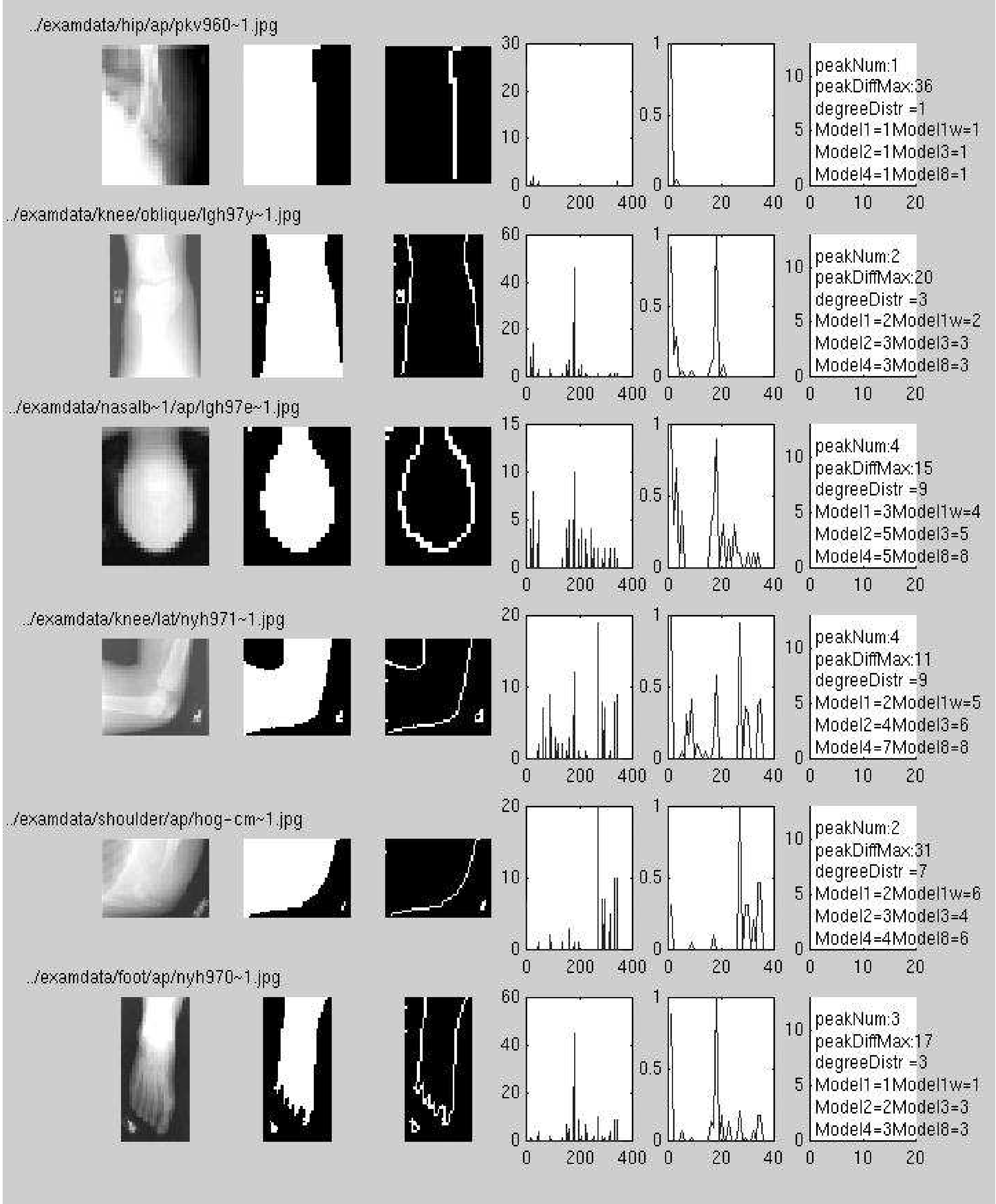


Fig 4-5: Typical shape images from the medical image database and their edge direction histograms

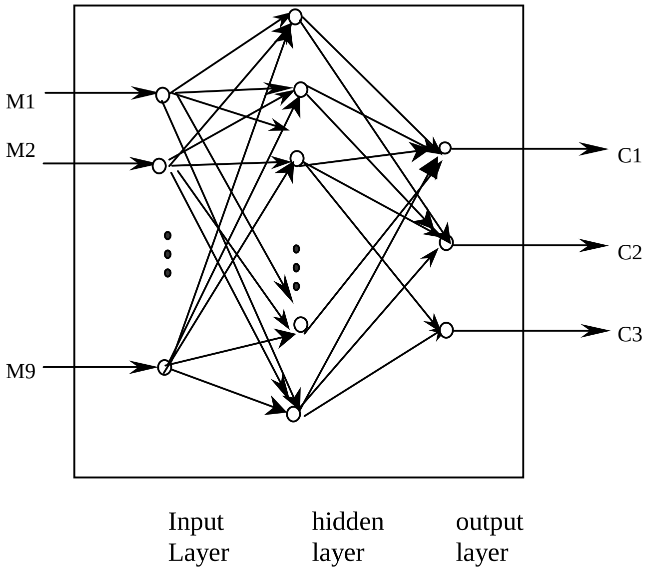


Fig 4-6: structure of the proposed neural networks

propose a multi-layer feedforward neural network to classify the main object shape. There are nine input nodes in the neural network corresponding to the nine features detected from the edge direction histograms, and three output nodes in the output layer with respect to the three classification classes. The decision rule is to assign an image to a class corresponding to one output node with the highest value. However, in the situation when two output nodes have very similar high values and it is hard to determine its classification, we assume the input image belongs to two classes, and the further verification in Section 4.3.4 will be used to clarify the ambiguity. Here we choose one hidden layer with 14 nodes, since our experiment results show that only one hidden layer is good enough to discriminate the feature data. Fig 4-6 shows the structure of the proposed neural network.

To train the neural network and find the weights, we use the back propagation-learning algorithm. Once the network is trained, it can be used for the classification.

4.3.4 Object Shape Detection

Although the edge direction features give a concise description of the shape, it still has a limited discrimination capability. For the further classification, we have to take use of the real shape information of the main object in the processed image. Thus, the goal of this step is to extract the entire shape contour of the main object and verify the classification with the extracted shape contour.

Since the edge map of a binary image obtained in Section 4.3.1 directly outlines the coarse shape contour of the main object in the image, we use it as the initial contour of the object. Then by using the proposed snake model in Chapter 3, we can detect a finer contour of the object. In order to detect the contour as accurate as possible, all processes in this step are performed on the original images. The obtained contour provides very important information of the processed image, such as the size, location and orientation of the main object in the processed image. This information will be used to adjust the knowledge model for a more detail content identification in section 4.4.

Another process performed in this step is to verify the classification results from the neural network. As we known, the nine shape features extracted from the edge direction histogram only give compact descriptions of the object shape. It is quite possible that two or more different shape may share very similar shape features. This will introduce errors in the classification. To solve this problem, we verify each classification result by comparing the detected object shape in the processed image with the global shape defined in its related knowledge model. If they are quite similar, we keep the original classifica-

ankle

elbow

Fig 4-7: The ankle and leg images

tion result; otherwise, we reject it.

4.4. Detail Content Identification

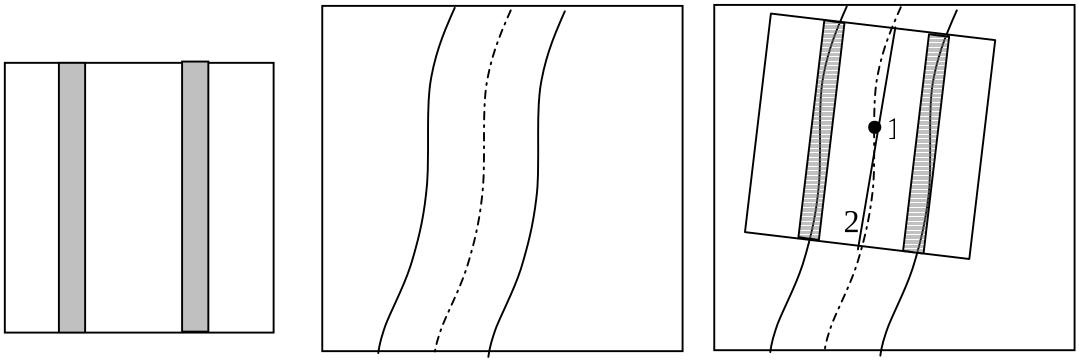
Using a single shape outline attribute for classification may lack sufficient discriminatory information. For example, two different body parts, such as ankle and elbow, may have very similar shape outlines as shown in Fig 4-7, but they do have totally different bone structures. To further classification, it is necessary to integrate the information of component objects in the knowledge model for a more elaborate matching. Thus in this finer classification stage, we emphasize the match of model component objects and their spatial relationships. The processed matching scheme consists of three steps: the first step tries to modify the model object shapes as close as the image objects and seek the most appropriate pose to locate the model in the images. The second step is to search for the shape feature of each component object in the related image region. Once all component objects have been found in the image, the third step is to perform the match of their spatial relations with the semantic graph defined in the knowledge model. The goodness of the matching is determined by a similarity measurement which depends on how many model component objects have been found in the images and how well their spatial relations satisfy the semantic graph defined in the model.

4.4.1 Modification of Models

As we mentioned earlier, the large variability of anatomical structures may influence the accuracy of model matching. Also the difference of view points, when radiographs are taken, may introduce the difficulties of appropriately locating the model in images. To solve these problems, we have to scale the model size as close as the real size of image objects, rotate it to correspond the image object orientation, and search for the most possible position to locate the model in images.

The size information is implied in the contour of the main object in the processed images. To obtain it, we take use of the contour detected by the snake model in Section 3.4.2, and extract its medial axis by a medial axis transform [Jeng 92]. With the help of the medial axis of the interested object, the size of the object can be easily estimated. Then with the reference of the knowledge model size, a scale global ratio r_g can be derived, which is used to scale the model size.

Regarding to the orientation, the medial axis of the main object actually provides enough information about the image object orientation. However, due to the size difference between the scaled model and the processed image (in most cases, the scaled model is smaller than the processed image), the matching is usually performed on parts of the processed image along the medial axis of the main object in images. Thus, to acquire



(a) The knowledge model (b) The processed image (c) Scale and rotate the knowledge model to match the processed image. 1: the model center 2: the fitted medial axis line

Fig 4.8: Illustration of matching a knowledge model to an image

the accurate orientation information, we only concentrate on the medial axis of the main object in the overlap part with the scaled model image and use a line to fit the medial axis. The orientation of the fitted line will be treated as the orientation of the image object, and directly used to rotate the knowledge model, as shown in Fig 4-8 (c).

As for the model position, we define it as the coordinate of model center, and constraint its movement only along the medial axis of the main object during the model matching.

The above size and orientation information gives a coarse description of the image object situation. We can use them as the initial setting to adjust the model. But to further accurately locate the model into the processed images, we try to seek the size parameter *r*, the position parameter *p*, and the orientation angle α of the model by using a maximum a *posteriori* approach.

$$\left\langle r_{MAP}\ , p_{MAP}\ , \alpha_{MAP} \right\rangle = \arg\max_{r,p,\alpha}\ P(r, p, \alpha \mid \nabla I) \qquad (4-10)$$

In this equation, ∇I is the gradient of the processed image. As the location and pose of the model are completely determined by *r, p* and *α*, this definition is helpful to search for a set of most appropriate *r, p* and *α*, given the current image. The maximum *a posteriori* can be computed from Equation (4-11) using Bayes' rule.

$$\begin{aligned} P(r,p,\alpha \mid \nabla I) &= \frac{P(\nabla I \mid r,p,\alpha)P(r,p,\alpha)}{P(\nabla I)} \\ &= \frac{P(\nabla I \mid r,p,\alpha)P(r)P(p)P(\alpha)}{P(\nabla I)} \end{aligned} \qquad (4-11)$$

Note, in this proceeding step, we assume that the scale ratio *r*, the position *p* and the orientation *α* are independent from each other. This is reasonable in our study. For computation efficiency, we discard the normalization term $P(\nabla I)$ in the denominator as it does not depend on the scale, position, or orientation.

Gradient Term: The first term $P(\nabla I \mid r, p, \alpha)$ in Eq. (4-11) computes the probability of seeing certain image gradients given the current shape model. Consider the relationship between the shape model determined by *r, p, α* and the gradient ∇I. If the shape model correctly locates in the processed image, we expect to see more image gradients in it, as shown in Fig 4-8 (c). More precisely, the closer between the image object shape and the model shape, the more image gradients correspond to the model edges. In order to make the model robust, we relax its shape constraint by extending the model shape edges using the relax matrix ***M*** in Section 2.4.2. Let $\boldsymbol{S} = \{(x_{si},y_{si})\}\ i=1..n$ be a vector describing the points on the shape edges of the relaxed model. We define the current model situation S^* computed from *r, p* and *α* as:

$$S^{*} = r\begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} S + \begin{bmatrix} p_x \\ p_y \end{bmatrix} \qquad (4-12)$$

The best fit pose of the model may be found by maximize the gradient probability, which is modeled as:

$$P(\nabla I \mid r, p, \alpha) = \frac{\sum_{x=1}^{N}\sum_{y=1}^{M} S^{*}(x,y) \bullet \mid \nabla I(x,y) \mid^{2}}{\sum_{x=1}^{N}\sum_{y=1}^{M} S^{*}(x,y)} \qquad (4-13)$$

Scale Ratio Priori: The second term *P(r)* in Eq. (4-11) is the scale ratio priori. As we described above, a global scale ratio can be derived from the model and the image object size information. However, because of the variability of anatomical structures, the global scale ratio sometimes doesn't represent the real object size precisely. From our experience, searching around the global shape ratio r_g is quite possible to find a best fit shape ratio r. Thus the scale ratio priori can be modeled as a gaussian distribution over the global shape ratio r_g with variance σ_r.

$$P(r) = \frac{1}{\sqrt{2\pi}\sigma_r} \exp\left(-\frac{(r - r_g)^2}{2\sigma_r^2}\right) \qquad (4-14)$$

Pose Priori: For the position priori *P(p)*, we simply assume a uniform distribution over all positions in the image object.

$$P(p) = U(-\infty, \infty) \qquad (4-15)$$

This is reasonable, since the expected image object can appeare at anywhere in the image, depending on how the radiologist takes the radiographs. In order to efficiently find the most appropriate position, we incorporate heuristics in the processing, which is that the model shape can be only found within the image object. So, we constraint the model movement such that its center have to move along the medial axis of the image object to seek the best fit position.

Rotation Priori: The last term *P(α)*, the rotation priori, in Eq. (4-11) employs the medial axis of the image object. According to the orientation of the fitted medial axis line in the overlap region, the model rotate itself until its medial axis align with medial axis the line. However, it is possible that the orientation of the medial axis line sometimes doesn't correspond to the image object orientation very well. In such situation, a little bit adjustment around the straight line would be very helpful for finding the best fit orientation. So the corresponding priori is designed as a gaussian distribution over the rotation range.

$$P(\alpha) = \frac{1}{\sqrt{2\pi}\sigma_\alpha} \exp\left(-\frac{(\alpha - \alpha_g)^2}{2\sigma_\alpha^2}\right) \qquad (4-16)$$

These terms define the maximum *a posteriori* estimator of scale, pose, and orientation, which estimates the final situation of the model. For efficiency, these three parameters are quantified in the processes. The final maximum position is used to locate the model in images.

4.4.2 Match of Models

Once a knowledge model is correctly located in the processed images, we extract a corresponding search image region for each component object, and then perform a correlation match between the component object shape and image object shape, as mentioned in section 2.4.2. A component object’s shape is assumed to be found in the processed image only if its correlation value exceeds a pre-defined threshold which is derived from a number of experiments. Once the model’s entire component object shapes are found in the processed image, their spatial relations will be checked by using the semantic graph match algorithm described in section 2.4.2. Our assumption used here is if the model finds all its component objects in the processed image and their spatial relations satisfy with the pre-defined spatial constraints, we say this model matches the processed image.

4.5 Evaluation of Match

Evaluation of a match result is another important issue in the system. In this section, we discuss two topics related to the evaluation: one is how to estimate the match accuracy between a knowledge model and a given image. In practical application, it is quite possible to have more than one knowledge model matching the processed image. Thus, the other topic to discuss is how to analyze the characteristics of the match results and use them to reject those unreasonable matches.

4.5.1 Confidence Level of Match

When evaluating a matched model, we are interested in how accurate the model can represents the processed image. According to the hierarchical match strategy of the knowledge model, the match accuracy should be based on not only the match of its component object shapes, but also the match of its semantic graph. So the matching confidence level can be defined as:

$$\mathit{Confidence\ \ Level}\ (m) = \begin{cases} - & \exists c_{mj} < th \quad (1) \\ 0 & \exists \mu_{mj} = 0 \quad \forall c_{mj} \ge th \quad (2) \\ \dfrac{\sum_{j=0}^{n} \mu_{mj} \cdot c_{mj}}{\sum_{j=0}^{n} \mu_{mj}} & \forall \mu_{mj} > 0 \quad \forall c_{mj} \ge th \quad (3) \end{cases} \qquad (4-17)$$

Where, m is the model identity, n is the total number of the component objects in the model. c_{mj} represents the match result of the j-th component object in the model m, and μ_{mj} describes the fuzzy membership of a match position of the j-th component object in the model m. If any one of model component object shapes in the given model is unable to be found in the processed image, we declare this knowledge model fails to represent the processed image and reject it, as presented in Eq. 4-17 (1). If all model's component object shapes are found in the processed images, but their spatial relations don't satisfy with the model defined spatial constraints, which means at least one of them lies outside of the pre-defined spatial position, we still think this model is not enough to precisely represent the image and set its confidence level to be zero, as shown in Eq. 4-17 (2). Only when the model finds all its component objects in the processed image and their spatial relations satisfy with the pre-defined spatial constraints, we say this model can be used to explain the processed image and set its confidence level as Eq. 4-17 (3). By using this confidence level, we have a quantitative measurement of the matching between a knowledge model and the processed image.

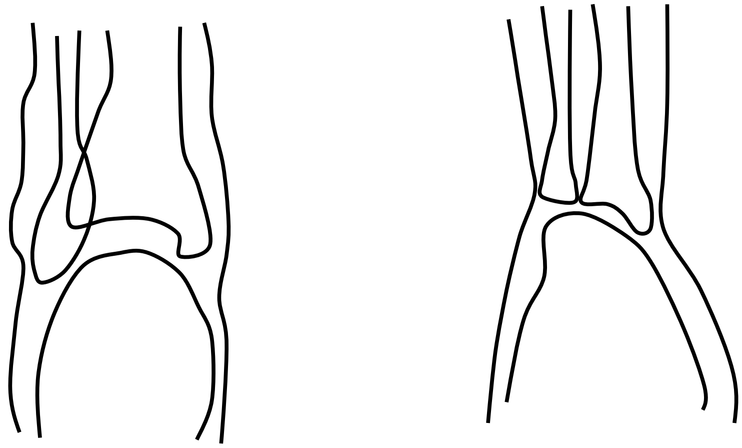
According to our hypotheses, only when the processed image features detected by a knowledge model satisfy with the model defined constraints, its confidence level can be greater than 0. In such situation, we assume the model succeed in the match procedure and classify the processed image into the image category represented by this model.

4.5.2 Evaluation Rules

During the classification, we use a set of knowledge models to understand the processed image. Ideally, the system is be able to find one model fit each input image and the rest models should be failed. But unfortunately, because of the similarity of anatomical structures of some human body parts, the corresponding knowledge models may be very close. It is quite possible that more than one model can succeed in matching the processed image. For example, the ankle model and wrist model, as shown in Fig 4.9 (a) and (b), all have three component objects and their spatial relations are nearly the same. So how to derive reasonable matches from these match models is what we want to solve in this step.

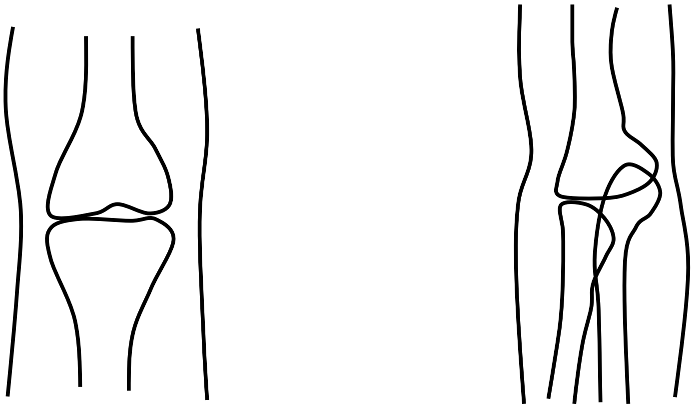
Our solution is to derive a set of evaluation rules based on the characteristics of the correct match results. The reason is that: the images belonging to the same image category must share some common features, such as the value range of the confidence level related to the correct knowledge model, and their processed results of other knowledge models. To explain it more clearly, we use the knee images as an example. After the coarse shape classification, the knee images are classified into cluster 2 because of their shape outline. In this cluster, there are other models sharing the same shape outline, such as the ankle, knee, wrist, pelvis, and elbow model as shown in Fig 4.9.

Based on our assumptions, all these models have to be used to match the processed



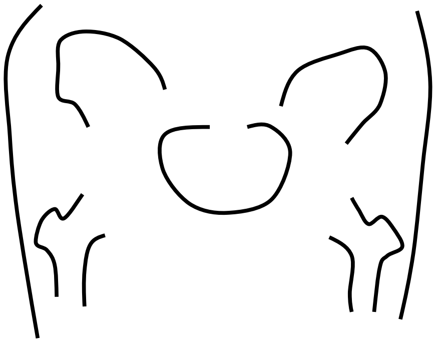
(a) The ankle model

(b) The wrist model



(c) The knee model

(d) The elbow model



(e) The pelvis model

Fig 4-9: The knowledge models

	Ankle	Elbow	Knee	wrist	pelvis
1. cedsi9_2.jpg	--	--	0.3516	--	--
2. lg5703_1.jpg	--	--	0.2971	--	--
3. lgd684_1.jpg	--	--	0.4971	1.5452	--
4. lgdb06_1.jpg	--	--	0.1703	--	--
5. lgdd06_1.jpg	0	--	0.2244	--	--
6. lgh97e_2.jpg	--	--	0.4413	--	--
7. pkv960_2.jpg	0	--	0.2322	--	--
8. qoa960_1.jpg	0	--	0.3772	0	--
9. qoa961_1.jpg	--	--	0.2524	--	--
10.qoa960_4.jpg	--	0	0.4285	--	--
11.qoa961_2.jpg	--	--	0.2999	--	--
12.qoa961_3.jpg	--	--	0.4350	--	--
13. cedsi9_1.jpg	--	--	0.2118	--	--
14. lg5804_1.jpg	--	--	0.1587	--	--
15. qoa960_2.jpg	1.6930	--	0.3368	0	--
16. lgh97e_1.jpg	--	--	0.1006	--	--
17. lgh97y_1.jpg	--	--	0.2858	--	--
18. lgh97y_2.jpg	--	--	0.2325	--	--
19. qoa960_3.jpg	0.9615	--	0.4864	0	--
20. pkv960_3.jpg	--	--	0.2966	--	--
21. qoa961_1.jpg	--	--	0.2524	--	--
22. qoa960_2.jpg	--	--	0.5534	--	--
23. qoa961_1.jpg	1.3447	--	0.3962	--	--
24. qoa961_11.jpg	0	--	0.1508	--	--
25. ucsf97_1.jpg	0	--	0.3273	--	--

Table 4-1: The match performance of different knee images
--: No match between the model and the image

images. Table 4-1 lists the confidence level of the match between the knee images and these knowledge models. The table clearly illustrates that the knee model works well on all input knee images, as shown in the knee column. This verifies our hypotheses. However, Due to the ambiguity of the input images, it may find other models which also work fine on the processed images, such as the wrist model on image 3, and the ankle model on the other three images (image 15,19,23). We like to think them as errors, and try to eliminate during reevaluation. So the following rules are derived to verify the knee images:

Rule 1: *If the processed image is a knee image, only knee model has the match results and all other model should be failed*

Rule 2: *If the processed image is a knee image, its confidence level should be within a certain range derived from most knee images.*

Based on the above rules, we can reevaluate the match results by checking if they satisfy with the corresponding rules and reject those violated match results. Table 4-2 shows the match results of another image category which corresponds to ankle images. It may note that every image shows the match confidence value in the knee column. That is because the knee model has a very simple structure which is possibly found as component structures in many other knowledge models. To avoid such situation, we use the above rules to reevaluate the match results. Table 4-3 demonstrates the results after the reevaluation. It is obvious that all match results in the knee column are rejected since they don't satisfy with the derived rules. As for the wrist image, we have a rule like:

Rule 3: *If the processed image is a wrist image, the match confidence level should be greater than 1.*

	Ankle	Elbow	Knee	wrist	pelvis
1. lg65c4_1.jpg	2.2357	--	0.4310	0	--
2. lg67c3_1.jpg	2.1725	--	0.9811	1.4295	--
3. lga6a4_1.jpg	2.1368	--	0.6431	--	--
4. qoa960_1.jpg	1.6056	--	0.3829	0.8047	--
5. qoa960_2.jpg	1.9908	--	0.4168	0	--
6. ucsf97_1.jpg	1.2476	0	1.2137	0	--
7.wh-uk9_1.jpg	1.1321	--	0.7389	0	--
8.wh-uk9_4.jpg	2.0529	0	0.7468	1.2214	--
9.wh-uk9_2.jpg	1.9208	--	0.7468	--	--
10.bjc960_1.jpg	2.2868	0	0.4522	0	--
11.lg2c45_1.jpg	1.8490	--	0.3193	0.9026	--
12.hog970_1.jpg	2.2354	--	0.2937	1.8820	--
13.lgaa43_1.jpg	1.5903	0	1.0610	0	--
14.lge8c4_1.jpg	2.0387	0	0.5739	0	--
15.lgh960_1.jpg	2.9603	--	0.4681	0	--
16.lgh970_2.jpg	2.0024	0	0.2896	--	--
17.lgh97e_1.jpg	1.4186	--	0.5278	1.6118	--
18.lgh97e_2.jpg	1.9669	--	0.3707	0	--
19.lgh97e_3.jpg	2.8880	--	0.4704	0.6674	--
20.lgh97e_4.jpg	2.0715	--	0.4318	1.2205	--
21.mer961_1.jpg	2.0087	--	0.7525	--	--
22.lgh97y_1.jpg	2.2920	0	0.5032	0.8981	--
23.lgh980_1.jpg	2.0054	--	1.0256	0.7136	--
24.qoa961_1.jpg	3.1762	0	0.7366	1.6622	--
25.wh-uk9_2.jpg	1.9208	--	0.2344	--	--

Table 4-2: The match performance of different ankle images
--: No match between the model and the image

	Ankle	Elbow	Knee	wrist	pelvis
1. lg65c4_1.jpg	2.2357	--	*	0	--
2. lg67c3_1.jpg	2.1725	--	*	1.4295	--
3. lga6a4_1.jpg	2.1368	--	*	--	--
4. qoa960_1.jpg	1.6056	--	*	*	--
5. qoa960_2.jpg	1.9908	--	*	0	--
6. ucsf97_1.jpg	1.2476	0	*	0	--
7.wh-uk9_1.jpg	1.1321	--	*	0	--
8.wh-uk9_4.jpg	2.0529	0	*	1.2214	--
9.wh-uk9_2.jpg	1.9208	--	*	--	--
10.bjc960_1.jpg	2.2868	0	*	0	--
11.lg2c45_1.jpg	1.8490	--	*	*	--
12.hog970_1.jpg	2.2354	--	*	1.8820	--
13.lgaa43_1.jpg	1.5903	0	*	0	--
14.lge8c4_1.jpg	2.0387	0	*	0	--
15.lgh960_1.jpg	2.9603	--	*	0	--
16.lgh970_2.jpg	2.0024	0	*	--	--
17.lgh97e_1.jpg	1.4186	--	*	1.6118	--
18.lgh97e_2.jpg	1.9669	--	*	0	--
19.lgh97e_3.jpg	2.8880	--	*	*	--
20.lgh97e_4.jpg	2.0715	--	*	1.2205	--
21.mer961_1.jpg	2.0087	--	*	--	--
22.lgh97y_1.jpg	2.2920	0	*	*	--
23.lgh980_1.jpg	2.0054	--	*	*	--
24.qoa961_1.jpg	3.1762	0	*	1.6622	--
25.wh-uk9_2.jpg	1.9208	--	*	--	--

Table 4-3: the reevaluation results of ankle images

* : The rejected result

--: No match between the model and the image

These rules can help us reject some impossible results. Once all rules are checked, the finial results can be used to classify the image.

4.6 System Evaluation and Experimental Results

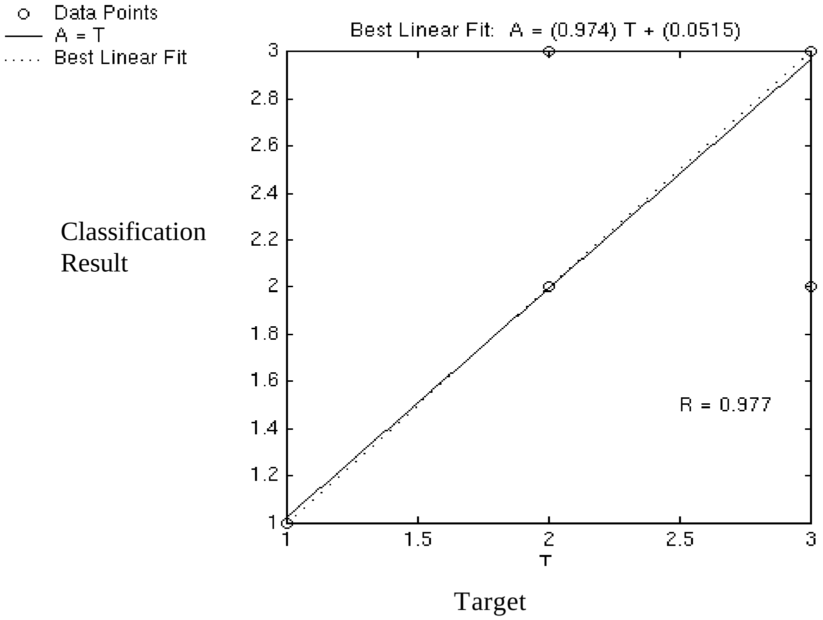
A medical image database of more than 400 images is used to test the performance of the content-based medical image classification prototype system. This medical image database contains a variety of human body parts radiographs, providing a rich source for testing. The presented system prototype is implemented using Matlab under SUN sparc 5 system with the main memory 250 MB.

4.6.1 Evaluation of Coarse Shape Classification

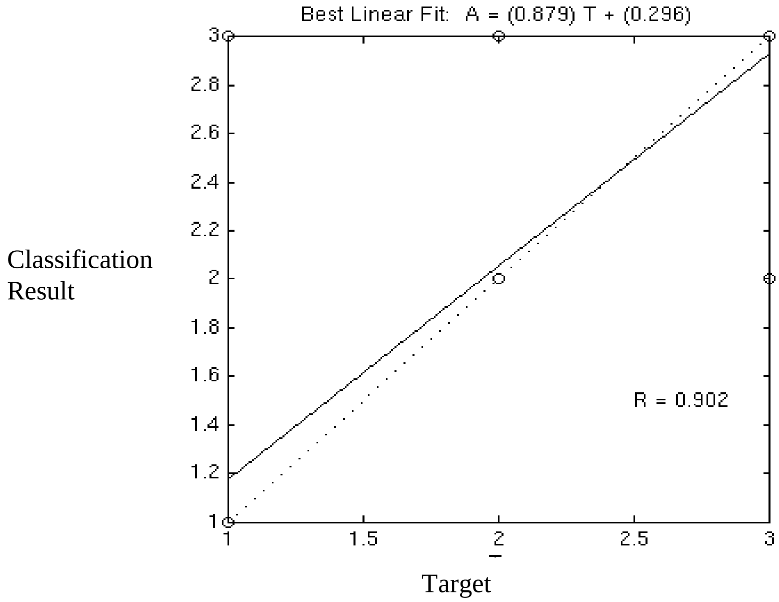
To study the performance of coarse shape classification, we randomly select 400 images from our medical image database. According to their visual shape outlines, we manually assign a class type to each of them and use that as the target value. Then these images are divided into two sets: the training set and the testing set. In the training phase, the neural network is trained 200 iterations of the training data. The trained classification neural network is then put through the testing data. The output decisions are compared with the target values. The overall correction percent is calculated by Eq. (4-18)

$$C = \frac{N_{correct}}{N_{total}} \times 100\% \qquad (4-18)$$

To measure the performance of a trained network in more detail, we also perform a regression analysis between the network response and the correspond targets. The



(a) The regression analysis of the training data



(b) The regression analysis of the testing data

Fig 4-10: The regression analysis of neural networks

N (times)	The training data set		The test data set	
	C	*R*	*C*	*R*
1	0.9732	0.977	0.9450	0.902
2	0.9464	0.953	0.9358	0.892
3	0.9375	0.946	0.9450	0.927
4	0.9732	0.977	0.9266	0.911
5	0.9908	0.992	0.8482	0.822
Average	0.9642	0.969	0.9201	0.8908

Table 4-4: The performance of the neural network

following figure, Fig 4-10, illustrates the graphical output of the regression analysis. A dashed line indicates the best linear fit. The perfect fit (output equal to targets) is indicated by the solid lines. *R*-value is a correlation coefficient between the classification output and target. It is a measurement of how well the variation in the output is explained by the targets. If its value is close to 1, we say a perfect correlation between targets and outputs.

To avoid the overfitting, we start from 5 different sets of random weights and train the neural network. Table 4-4 demonstrates the performance of the neural networks on the overall correction percent (*C)* and the correlation coefficient between the classification result and expected target (*R).* From the above experimental results, it is obvious that the overall correction percents of the test data set are around 90%. This is sufficient to demonstrates that the extracted nine scale, rotate and translate invariant features have enough discrimination ability to classify the input images into one of the three clusters.

Regarding to the processing efficiency, the time-consuming processes in the coarse shape classification are the image preprocessing, feature extraction, and contour detection.

As for the neural network classification, once the neural network is trained, the classification procedure is quite efficient.

4.6.2 Evaluation of Detail Content Classification

Based on the characteristics of human body structures, most radiographs in our medical image database can be classified into cluster 2 (the two peak cluster) such as leg, elbow, pelvis radiographs and etc. Thus, the detail content classification is mainly performed on the images belonging to this cluster.

To test the performance of detail content classification, we construct a database which includes 200 images selected from five image categories: Elbow radiographs, Ankle radiographs, Knee radiographs, Pelvis radiographs, Wrist radiographs. Table 4-5 list the performance of the system working on these images

Image categories	Image number	Process time/image	Number of image with Correct match	Correct percent
Elbow images	10	64 seconds	6	60%
Ankle images	50	58 seconds	36	72%
Pelvis images	40	80 seconds	30	75%
Knee images	50	45 seconds	41	82%
Wrist images	50	62 seconds	32	64%
total				70.6%

Table 4-5: the performance of detail content classification

The computation cost of detail content classification is determined by the number of knowledge models needed for the match and the complexity of knowledge model. For the five models in examples, the pelvis model is the most time-consuming one, as it has more component objects and its detection and match schemes are more complicated. Compared to the pelvis model, the rest knowledge models are more efficient. Table 4-5 also lists the process time of each knowledge model. Since all these tests are performed using Matlab, there is reason to believe that the performance of the system can be greatly improved if the whole process is programmed by C/C++. As the knowledge models are totally independent, another way to speed up the system is to design the match of each knowledge model parallel.

4.7 Summery

In this chapter, we have presented a content-based medical image classification system. The system employs both global and local features to classify the input images by performing a two-stage processing. The first stage, coarse shape classification, focuses on the global shape features of objects in the processed images. Through the coarse shape classification, the input image is classified into one of three global shape clusters. The purpose of this stage is to reduce the number of models needed for detail matching and speed up the processing. The experiments show that the first stage can be performed very fast. Once the coarse shape classification is done, a small set of plausible model class candidates are then presented for a detailed matching in the second stage. In this stage, local features used to identify salient component objects in images are employed and incorporated with their spatial relation to discard those spurious match candidates. The output match results imply a set of model which might be existed in the processed images. To further confirm them, an evaluation strategy is used to reject those unreasonable results and inference the most possible match models.

The experimental results suggest that the system prototype is applicable and robust, and its performance satisfies with our requirements.

Chapter 5:

Conclusion and Future work

5.1 Conclusion

In this dissertation, we discussed the design and implementation of a knowledge-based medical image classification system.

We began by designing a knowledge model for medical image content analysis based on the object-oriented paradigm. The new model abstracts common properties from different types of medical images by using three attributes: description, component, and semantic graph, and specifies its actions to schedule the detection procedure, properly deform the model shapes to match the corresponding anatomies in images, select the best match candidates, and verify the candidates' spatial relations with the semantic graph defined in the model. The advantages of the new model are: 1) It presents a robust representation ability to represent the object variability and relations, such as the spatial adjacency relation and the hierarchy relation. 2) It provides a tighter bond between features and methods. 3) It supports the complex image analysis and works on both normal and abnormal images.

Subsequently, we consider the problem of segmenting the object of interest from images. We proposed a region model to combine both region and edge information together in the segmentation, which directly overcome the problems associated with the initialization. Based on the proposed region model, we reformulate the traditional snake model and level set models. As expected, the performance of the reformulated models is greatly improved and experimental results are encouraging.

Finally, we integrate the above techniques to design the knowledge-based image classification system. The system employs both global and local features to classify the input images by performing a two-stage processing. The first stage, coarse shape classification, focuses on the global shape features of objects in the processed images. Through the coarse shape classification, the input image is classified into one of three global shape clusters. The purpose of this stage is to reduce the number of models needed for the detail matching and speed up the processing. Once the coarse shape classification is done, a small set of plausible model class candidates are then presented for a detailed matching in the second stage. In this stage, the local features used to identify salient component objects in images are employed and incorporated with the component objects' spatial relation to discard those spurious match candidates. The output match results imply a set of models which might be existed in the processed images. To further confirm them, an evaluation strategy is used to reject those unreasonable results and inference the most possible match models.

The performance of the system has been tested on our image databases, which contains more than 400 computer radiographs with different complexity body parts such as chest, pelvis, ankle, elbow and etc. The experimental results suggest that the system prototype is applicable and robust, and the accuracy of the system is nearly 70% in our image databases.

5.2 Future work

In this section, we present two general research directions which can be pursued in the future: Statistical analysis of anatomical shapes and medical image indexing.

5.2.1 Statistical Analysis of Anatomical Shapes

A natural extension of the current work is to consider the statistical shape analysis of anatomical structures in medical images.

In the domain of medical application, the capture of object variability is especially important. A shape model for a given anatomical object should be able to resemble a large portion of the shape instances that can be found either in different individual (interpatient variability) or within the same individual (intrapatient variability). This effect can be achieved by a statistical analysis of a set of shape instances of the object of interest. The parameter vector of the shape parameterization is interpreted as a random vector. The parameter vectors of a set of shape instances define a multivariate distribution that can be described by the infinite series of moments of the random vector. If the distribution is Gaussian, it is already determined by the first (mean) and second (covariance) moments. In the case of a non-Gaussian distribution, the first and second moments can still be used as an approximation. Thus, in the Gaussian approximation, the shape variability is captured by the covariance matrix of shape vectors representing a set of objects. The eigenvectors of the covariance matrix can be interpreted as fundamental modes of variation, or short eigenmodes. A subset of these modes can be used as a set of basis vectors defining an approximated shape-space of the set of objects. A weighted sum of these basis vectors and the mean shape is well suited for the generation of new shape samples in order to support tasks such as shape recognition or image segmentation. The appealing property of this approach is the ability to create modes that depend on a small number of parameters (a small number of eigenmodes), even for intricately shaped objects. Among different types of shape representations and parameterizations used for statistical shape models, Point distribution models [Cootes 92] with parameter vectors being composed of the coordinates of contour points have been frequently applied. A crucial requirement for this method is the one-to-one correspondence between the elements of parameter vectors for different shape instances. All vector elements with the same index must represent corresponding shape information.

5.2.2 Medical Image Indexing

Another problem to be addressed in medical image management is indexing, the assignment to each document of a synthetic descriptor facilitating its retrieval. Medical image index strategies, to date, have been primarily based on a few textual or numerical fields residing in standard relational database structures. Nearly all indexing and retrieval operations have been based on the patient identity, date, types of examination or image number. There is reason to believe that better understanding of disease processes might become possible if image databases can be organized and retrieved based on image content.

Our anatomical knowledge model provides an efficient way to describe the contents in medical images and organize them in an object hierarchy. Based on it, a new hierarchical index scheme can be generated to represent a complex image. The advantage of this feature index structure is that it includes not only feature parameters which identify variations between the image objects and their related model objects but also the hierarchical and spatial relations of objects in the image.

Reference:

[Amini et. al. 90] A. A. Amini, T.E. Weymouth and R.C. Jain, Using Dynamic Programming for Solving Variational Problems in Vision. *IEEE Transactions on Pattern Analysis & Machine Intelligence*. vol 12, n 9,pp. 855-867, 1990

[Ashka and Modestino 78] G.P. Ashkar and J.W. Modestino, The Contour Extraction Problem with Biomedical Applications. Computer Graphics Image processing. vol 7, pp 331-355, 1978

[Aggarwal and Cai 99] Aggarwal, J K. Cai, Q. Human motion analysis: A review *Computer Vision & Image Understanding.* v 73 n 3 pp 428-440, Mar 1999.

[Bach et. al 96] J.R. Bach, C. Fuller, A.Gupta, Arun Hampapur, et al. The Virage image search engine: An open framework for image management. *Proc. SPIE Storage and Retrieval for image and Video Database*, vol 2670, pp. 76-87,1996.

[Berger and Mohr 90] M.O.Berger and R.Mohr, Towards autonomy in active contour models, *IEEE 10th International Conference on Pattern Recognition*. pp. 847-851, 1990

[Brook 81] R.A.Brooks, Symbolic reasoning among 3-D models and 2-D images. *Artificial Intelligence*. vol 17, pp285-349, 1981

[Binford 82] T.O.Binford, Survey of Model-based Image Analysis System. *International Journal of Robot. Res*. vol 1, n 1, pp. 18-64, 198

[Canny 86] J.Canny, A Computational Approach to Edge Detection. *IEEE Trans. Pat. Anal. Mach. Intell.,* vol PAMI-8, pp. 679-698, 1986

[Chakraborty et. al. 96] A. Chakraboty, H. staib, and J. Duncan, Deformable Boundary Finding in Medical Images by integrating Gradient and Region Information. *IEEE Trans. on Medical Imaging*, vol 15, n 6, pp. 859-870, 1996

[Clouard et al 93] R. Clouard, C. Porquet, A. Elmoataz and M. Revenu, Resolution of image processing problems by dynamic planing within the framework of the blackboard model, *Intelligent Robots and Computer Vision XII, SPIE* 2056, pp. 419-429, 1993

[Cohen 91] L.D. Cohen, Note: On active Contour Models and Balloons. *CVGIP: Image Understanding*.Vol.53, no.2, pp211-218, 1991.

[Cohen and Cohen 93] L. D. Cohen, I. Cohen, Finite-element methods for active contour models and balloons for 2-D and 3-D images. *IEEE Transactions on Pattern Analysis & Machine Intelligence*. vol 15 n 11,pp. 1131-1147, Nov 1993

[Cohen and Kimmel 97] L. D. Cohen, R. Kimmel, Global minimum for active contour models: A minimal path approach. *International Journal of Computer Vision*. vol 24, n 1, pp. 57-78, 1997.

[Cootes et. al 92] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J.Graham. Training models of shape from sets of examples. *Proceedings of the British Machine Vision Conference*. vol 1, pp. 9-18, Springer Verlag, 1992

[Cox et al 96] I. J. Cox, S. B. Rao, Y, Zhong, “Ratio regions” a technique for image segmentation. *Proc, Int. Conf. Pattern Regonition*, pp 557-564, 1996

[Crevier and Lepage 97] D. Crevier and R. Lepage, Knowledge-based Image Understanding System: A Survey. *Computer Vision and Image Understanding*. vol 67, No 2, pp. 161-185,1997

[Draper et al 89] B.A.Draper, A.R.Hanson, and E.M.Riseman, The Schema System. *International Journal of Computer Vision*. vol 2, pp.209-250, 1989

[Draper et. al. 96] B.A. Draper, A.Hansen, E.Riseman, Knowledge-directed vision: Control, learning, and intergration. *Proceedings of IEEE*, vol. 84, No. 11 1996

[Davatzikos and Prince 95] C. A. Davatzikos, J. L. Prince, Active contour model for mapping the cortex. *IEEE Transactions on Medical Imaging*. vol 14, n 1, pp. 65-80,1995.

[Fua and Brechbuhler 97] P. Fua, C.Brechbuhler, Imposing hard constraints on deformable models through optimization in orthogonal subspaces. *Computer Vision & Image Understanding*. vol 65 n 2 .p 148-162, Feb 1997.

[Gonzalez and Dankel 93] A.J. Gonzalez, D.D. Dankel, The Engineering of Knowledge-based system: Theory and Practice. Prentice Hall, New Jersey 1993

[Gunn and Nixon 93] S. R. Gunn, and M. S. Nixon, A Robust Snake Implementation: A Dual Active Contour. *IEEE Trans on Pattern Analysis and Machine Intelligence*. vol 15, no.11, pp.1131-1147, 1993

[Haralick 84] R.M. Haralick, Digital Step Edges from Zero Crossing of Second Directional Derivatives. *IEEE Trans. Pat. Anal. Mach. Intell.*, vol PAMI-6, pp. 56-68, 1984

[Hansen and Henderson 89] A.R.Hansen and T.C.Henderson, CAD-based computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol 11, n 11, pp 1181-1193, 1989

[Huang et. al 96] T.S. Huang, S. Mehrotra, and K. Ramchandran, Multimedia analysis and retrieval system (MARS) project. *Proc of 33rd Annual Clinic on Library Application of Data Processing Digital Image Access and Retrieval,* 1996

[Ikeuchi and Kanade 88] K. Ikeuchi, T. Kanade, Automatic generation of object recognition programs, *Proc. IEEE*, vol 76, n 8, pp. 1016-1035, 1988.

[Ivins and Porrill 94] J. Ivins and J. Porrill, Active Region Models for segmenting Textures and colors. *Image and Vision Computing*, vol.13, no.5, pp.431-438, 1994

[Jeng and Sahni 92] J.F.Jenq, S. Sahni, Serial and parallel algorithms for the medial axis transform *IEEE Transactions on Pattern Analysis & Machine Intelligence*. vol 14, n 12. pp 1218-1224, 1992

[Kaper et al 85] J.N.Kaper, P.K. Sahoo, A.K.C. Wong, A new method for gray level picture thresholding using the entropy of the histogram. *Computing Vision Graphics Image Processing*, vol 29, pp 273-285, 1985

[Kass et al 87] M. Kass, A. Witkin and D. Terzopoulos, Snakes: Active Contour Models. *Proceedings of First International Conference on Computer Vision*, pp. 259-269. 1987

[Karaolani et. al. 92] P.Karaolani, G.D.Sullivan, and K.D.Baker. Active contour using finite elements to control local scale. *Proceedings of the British Machine vision Conference*, pp 472-480, Springer Verlag. 1992

[Kosko 97] B. Kosko, Fuzzy Engineering. Prentice Hall, 1997

[Leroy et al 96] B. Leroy, I.Herlin, and L.D.Cohen, Finite-element methods for active contour models. *12th Int. Conf. Analysis and Optimization of Systems,* pp.58-65,1996

[Lai 94] K. F. Lai, Deformable Contours: Modeling, Extraction, Detection and Classification. *Ph.D thesis, Univeristy of Wisconsin-Madison*. 1994

[Li and Wang 99] X. Li, and J. Wang, Adaptive Balloon Models. *Int. Conf. Computer Vision and Image Processing*. pp434-439, 1999

[Lin 88] W.C.Lin, Y.T.Weng, C.T.Chen, Expert vision systems integrating image segmentation and recognition processes, *International Journal on Engineering Applications of AI*. vol 1, n 4, pp. 230-247, 1988

[Levine and Shaheen 81] M.D.Levine, S.I. Shaheen, A Modular Computer vision system for picture segmentation and interpretation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol 3, n 5, pp. 540-555, 1981

[Luo et. al 99] H. Luo, R. Acharya, R. Gaborski Fully Automatic Detection of structures in Chest Radiographs. *International ICSC Congress on Computation Intelligence: Methods Applications,Rochester,* New York June 1999

[Ma and Manjunath 97] W.Y. Ma and D.S. Manjunath, Netra: A toolbox for navigating large image databases. *Proc IEEE Int. Conf. On Image Proc.* 1997

[Matsuyama 89] T.Matsuyama, Expert system for image processing: Knowledge-based composition of image analysis processes, *Computer Vision, Graphics and Image Processing*. vol 48, pp. 22-49, 1989

[Malladi et. al. 95] R. Malladi, J. A. Sethian, B.C. Vemuri, Shape Modeling with Front Propagation: A Level Set Approach. *IEEE Transactions on Pattern Analysis & Machine Intelligence*. vol 17, n 2, pp. 158-174, 1995

[Martelli 72] A.Martelli, Edge Dection Using Heuristic Search Methods. Computer Graphics Image Processing, vol 1, pp. 169-182, 1972

[Marr and Hildreth 80] D. Marr and E. Hildreth, Theory of Edge Detection. *Proceedings of Royal Society of London*, B207, pp. 187-217, 1980

[McKeown and Harvey 87] D.M.McKeown, W.A.Harvey, Automatic knowledge acquisition for aerial image interpretation. *SPIE vol758 Image Understanding and the Man-machine Interface*, pp. 144-164, 1987

[Menet et. al. 90] S. Menet, P. Saint-Marc, G. Medioni, Active contour models: Overview, implementation and applications. *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics.* Piscataway, NJ, USA, pp. 194-199, 1990.

[Minsky 75] M.Minsky, A framework for representing knowledge. *The Psychology of Computer Vision*, McGraw-Hill, New York, 1975

[Neuenschwander et. al 94] W. Neuenschwander, P. Fua, G. Szekely, O. Kubler, Making snakes converge from minimal initialization. *IEEE Proceedings - International Conference on Pattern Recognition*. vol 1, pp. 613-615,1994.

[Nazuf and Levine 84] A.M.Nazif, M.D.Levine, Low level image segmentation: an expert system. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol 6, no 5, pp 555-577, 1984

[Niblack et al 94] W. Niblack, R. Barber, and et al, The QBIC project: Querying images by content using color, tecture, and shape. *Proc. SPIE Storage and Retrieval for Image and Video Databases*, Feb 1994

[Nishiyama et al 94] H. Nishiyama, S. Kin, T. Yokoyama, Y. Matsushita, An Image Retrieval System Considering Subjective Perception. *Proc 1994 ACM SIGCHI Conf*, Boston, MA, April 1994, pp30-36

[Orphanoudakis et al 94] S. C. Orphanoudakis, C. Chronaki, and S. Kostomanolakis, I^2C: A System for the Indexing, Storage, and Retrieval of Medical Images by Content. *J. Medical Informatics*, Vol. 19, no. 2, pp.109-122, 1994

[Paragios and Deriche 99] N. Paragios, R. Deriche, Geodesic active regions for supervised texture segmentation. *Proc. IEEE Int'l Conf. Computer Vision.*1999

[Pavlidis and Liow 90] T. Pavlidis, Y. Liow, Integrating region growing and edge detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol 12, pp 225-233, 1990

[Pentland et al 96] A. Pentland, R. W.Picard, and S. Sclaroff. Photobook: Content-based manipulation of image database. *International Journal of Computer Vision*, 1996

[Poggio and Torre 84] T. Poggio and V. Torre, Ill-posed Problems and Regularization Analysis in Early Vision, *Proceedings of AARPA Image Understanding Workshop*, pp. 257-263, 1984

[Ronfard 94] R.Ronfard, Region-based Strategies for active contour models. *Int'l J. Computer Vision*,vol.13 no.2,pp.229-251, 1994

[Smith and Chang 96] J. R. Smith, S. F. Chang: Visualseek: A fully automated content-based image query system. *Proc ACM Multimedia 96*, 1996

[Sullivan et. al 93] S. Sullivan, L. Sandford, J. Ponce, On Using Geometric Distance Fits to Estimate 3D Object Shape, Pose, and Deformation from Range, CT, and Video Images. *Proceedings of the Int. Conf. on Comp. Vision and Pattern Recogn.* (CVPR'93), pp 110-115, 1993.

[Tan et al 92] H.L.Tan, S.B. Gelfand and E.J. Delp, A Cost Minimization approach to Edge Detection Using Simulated Annealing. *IEEE Trans. Pat. Anal. Mach. Intell.*, vol PAMI-14, pp. 3-18, 1992

[Tikhonov 63] A.N. Tikhonov, Regularization of incorrectly posed problems, SOV. Math. DOKL, vol 4, pp 1624-1627, 1963

[Tsai et. al. 93] C. T. Tsai, Y. N. Sun, P. C. Chung, Minimising the energy of active contour model using a Hopfield network. *IEE Proceedings-E Computers & Digital Techniques*. vol 140, n 6, pp 297-303, 1993.

[Walker 90] E.L.Walker, Exploiting geometric relationships for object modeling and recognition. *Proceedings of the SPIE Conference on Intelligent Robots and Computer Vision IX: Neural, Biological and 3D Methods*, pp. 353-363, 1990

[Williams and Shah 92] D. J. Williams, M. Shah, Fast algorithm for active contours and curvature estimation. *CVGI: Image Understanding*. vol 55, n 1, pp. 14-26, Jan 1992

[Wong and Huang 97] S.T.C. Wong,. H K. Huang, Networked multimedia for medical imaging. *IEEE Multimedia*. vol 4, n 2, pp 24-35, 1997.

[Xu et al 94] G. Xu, E.Segawa, and S.Tsuji, Robust Active Contours with Insensitive Parameters, *Pattern Recognition*, vol. 27, no.7, pp879-884, 1994

[Xu and Prince 98] C.Xu, J.L.Prince, Snakes, Shapes, and Gradient Vector Flow, *IEEE. Trans Image Processing*, vol. 7, no. 3, 1998

[Zadeh 65] L.A.Zadeh, Fuzzy Sets, *Information and Control*, vol 8, pp338-353, 1965

[Zhu and Yuille 96] S. Zhu and A. Yuille, Region Competition: Unifying Snakes, Region Growing, and Bayes/ MDL for Multiband Image Segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol 18, pp 884-900, 1996