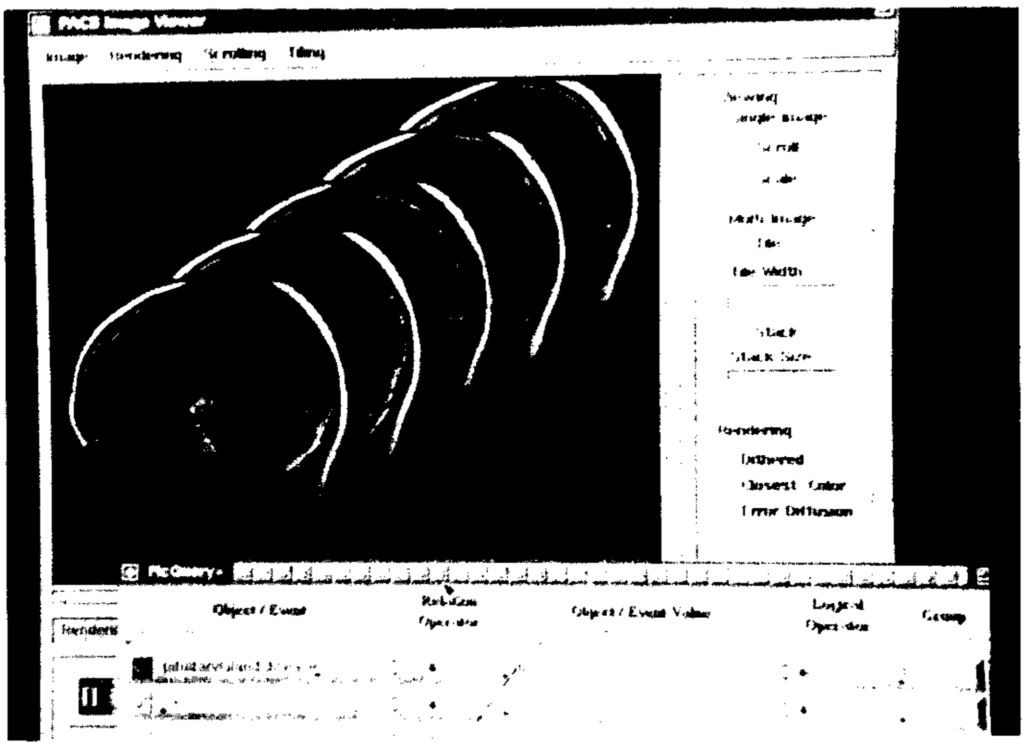
Computer Science Department Technical Report
University of California
Los Angeles, CA 90024-1596

A KNOWLEDGE-BASED MULTIMEDIA MEDICAL DISTRIBUTED DATABASE SYSTEM -- KMED

W. W. Chu
A. F. Cardenas
R. K. Taira

March 1993
CSD-930005

A Knowledge-Based Multimedia Medical Distributed Database System — KMeD



University of California
Los Angeles

A Knowledge-Based Multimedia Medical Distributed Database System — KMeD*

Wesley W. Chu and Alfonso F. Cardenas

UCLA Computer Science Department

and

Ricky K. Taira

UCLA Department of Radiological Sciences

SIGNIFICANCE

Medical research databases have several advanced requirements. Firstly, on-line access to the large repository of data stored in distributed patient information systems is needed. This data includes structured text (e.g., demographic information), free text with imprecise medical terms and descriptions (e.g., pathology reports), numerical data (e.g., cardiac monitoring), diagnostic images (e.g., computed tomography, magnetic resonance), and voice data (e.g., dictated radiology reports).

Secondly, medical research databases require the ability to query the collected information repository using not only traditional query keys (e.g., patient name, hospital ID, etc.) but also more sophisticated keys based on image content, evolutionary transformations of objects (e.g., organ development, disease progression, etc.), spatial relationships between objects, and temporal relationships between object events.

Thirdly, the most sophisticated systems require the computer to apply appropriate methods for analyzing and visualizing retrieved data in a manner mimicking that of trained medical specialists. This requires a knowledge of medical terms and available processing algorithms, as well as conceptual models of the anatomy, physiology, and pathology of an organ system of interest.

Currently a framework for an intelligent medical database system which can retrieve, analyze, and conveniently visualize distributed medical information using a high-level easy-to-use graphical interface is a major need.

* This work is supported by a National Science Foundation Scientific Database Initiative Grant No. IRI9116849.

PROJECT GOALS

The UCLA Knowledge-based Multimedia Medical Distributed Database System (KMeD) is being developed to help manage general medical research projects which involve understanding the correlation of patient symptoms, image features, and underlying disease processes for various patient profiles. In particular, we have initially focused on developing a database system for musculo-skeletal research (with future extensions to brain modeling research). We have collected digital hand radiographs from the radiology picture archiving and communication system (PACS) and have segmentation software to extract bone features [Pietk91]. Among the goals of the KMeD project are:

- Allow users to query the medical database by image content.
- Allow users to specify medical objects, e.g., bone growth, and evolution queries.
- Provide analysis and presentation methods for visualization of structural hand models for both normal and abnormal patient cases.
- Allow the user to formulate queries using imprecise medical terms.

USER REQUIREMENTS

Users of the KMeD system will be scientists, researchers, and clinicians unfamiliar with low-level database languages. The user should be able to easily perform the following:

- Specify the data domain space among a multimedia database federation,
- Graphically visualize underlying data models, knowledge hierarchies, and domain rules,
- Incorporate high-level abstract data types,
- Use imprecise (fuzzy) object descriptors and imprecise relational correlators to formulate query predicates,
- Process temporal and evolutionary queries,
- Specify analysis methods for retrieved multimedia data, and
- Specify visualization methods for retrieved and processed multimedia data results.

DATA MODELING

Accurate data models are of great interest in medical databases due to the complexity of object features and the complexity of object relationships. The requirement for storing, analyzing, and visualizing image data lead to the stacked image data model. The need to model the evolution and dynamics of objects in the body lead to the temporal evolutionary data model.

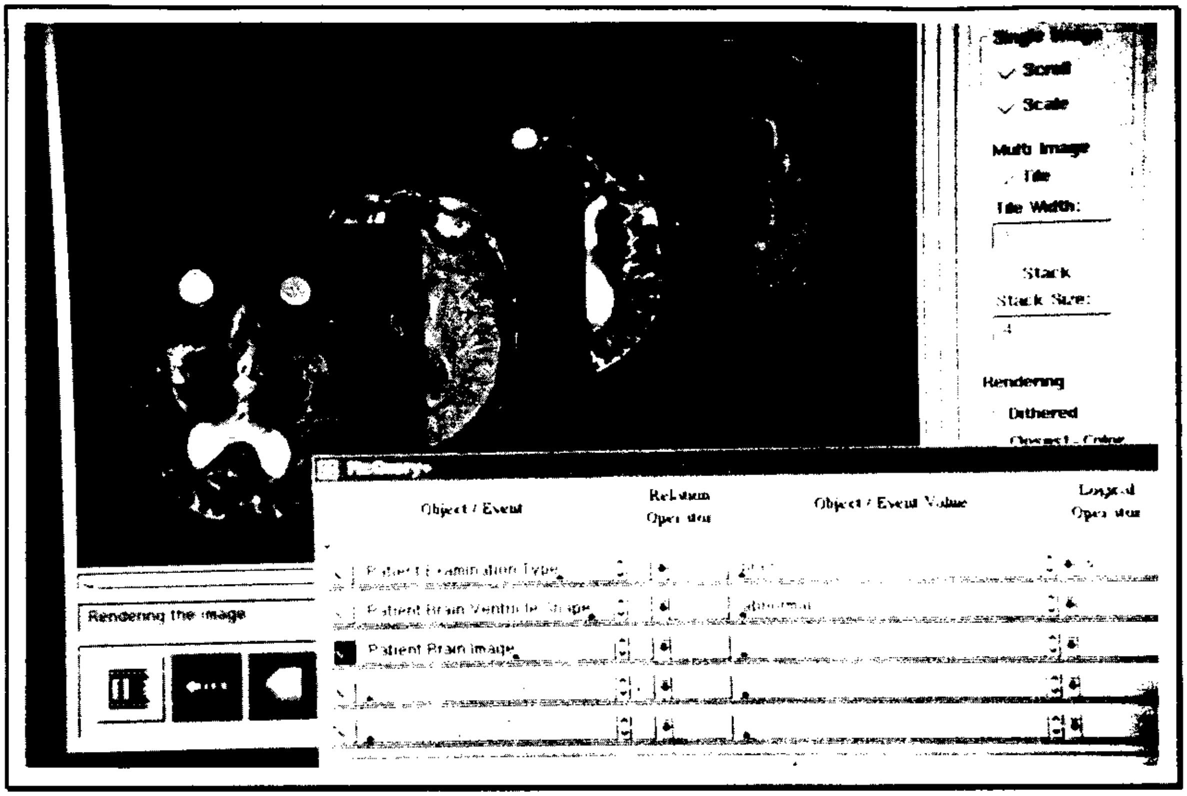


Fig. 1 - *Brain image stack with PICQUERY+ user interface table.*

Stacked-Image Data Model

Image data is modeled using a dynamic stacked logical data structure [Josep88]. Stacks consist of two-dimensional variables or pictures registered to the same gridded coordinate system. Multiple stacks of pictures can provide, for example, additional temporal, spectral, and/or spatial information. Stacks may consist of a hierarchy of other stacks. Furthermore, pictures composed of objects with pictorial as well as non-pictorial attributes and relationships can be represented. Thus, objects at different levels of abstraction can be modeled including time varying image stacks (see below). KMeD users may view an image stack at a high level while still maintaining the flexibility to manipulate images with pixel-level granularity. Figure 1 shows a magnetic resonance brain stack varying both spectrally and spatially.

Temporal Evolutionary Data Model

Our temporal evolutionary data model (TEDM) handles complex queries involving the comparison of the lifespan of two objects and/or events, the creation of objects, the fusion of objects, the fission of an object, and the gross transformation of object properties [Chu92a]. Formal constructs for the temporal relationships between objects, the specification of temporal constraints, and the inheritance characteristics passed between objects have been developed. TEDM uses evolutionary networks for modeling various

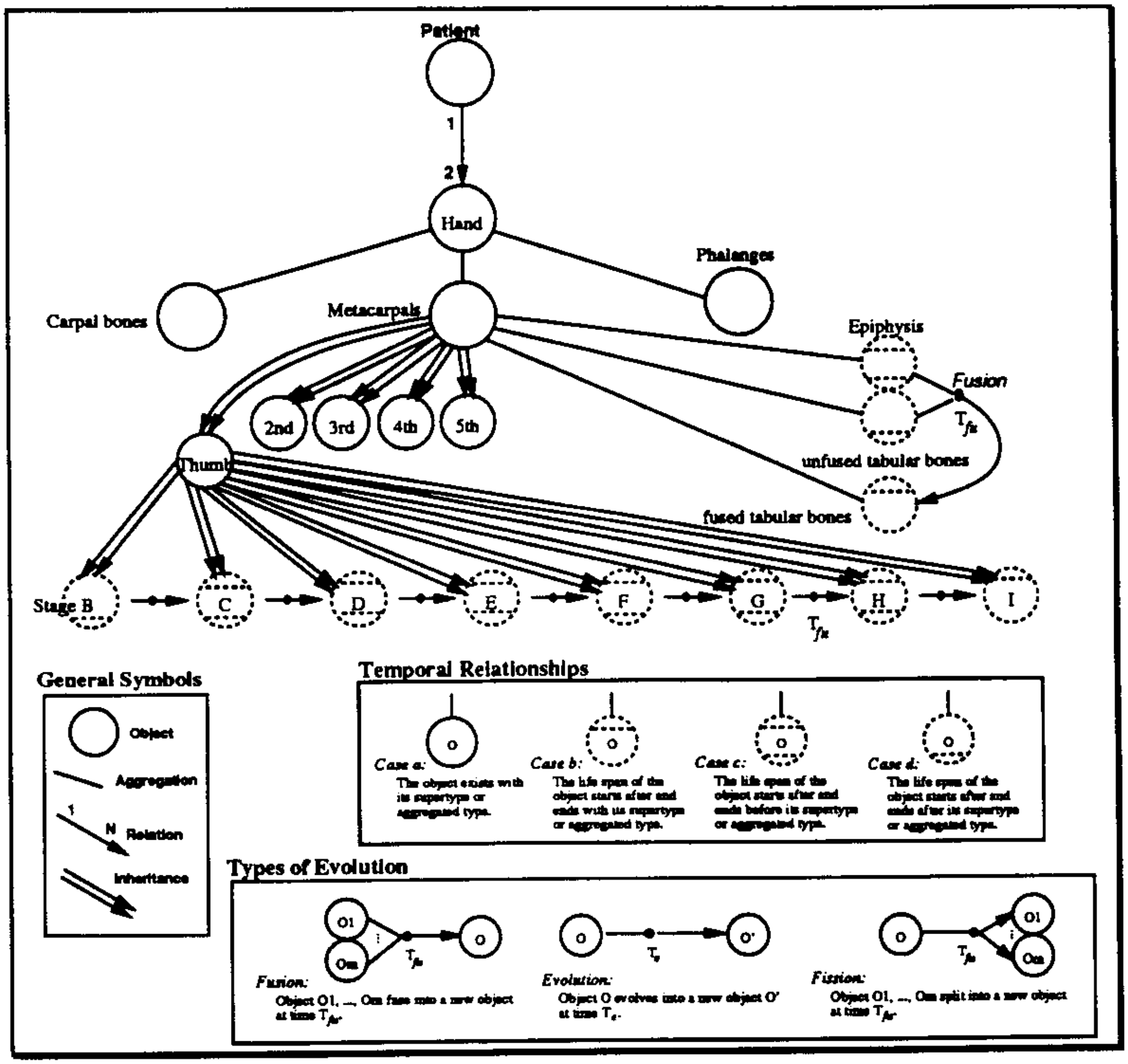


Fig. 2 - *TEDM for the object human hand. The model incorporates creation of new bones, fusion of existing bones, and the developmental evolution of existing bones.*

object transformation processes and includes inheritance rules between objects that exist in various time domains. Such a model allows us to store and retrieve images capturing temporal and evolutionary processes. Figure 2 shows the TEDM for the fusion and evolution of bones in the human hand.

◆ QUERY LANGUAGE AND USER INTERFACE

The high-level domain-independent query language PICQUERY$^+$ has been designed, and a subset of it is being implemented for the KMeD project [Carde93]. Queries are specified using either a tabular user interface (Figure 1) or in a more application specific user friendly graphical interface (Figure 3). Both interfaces allow the user to specify high level operations including query definition (including temporal and evolutionary predicates), data analysis

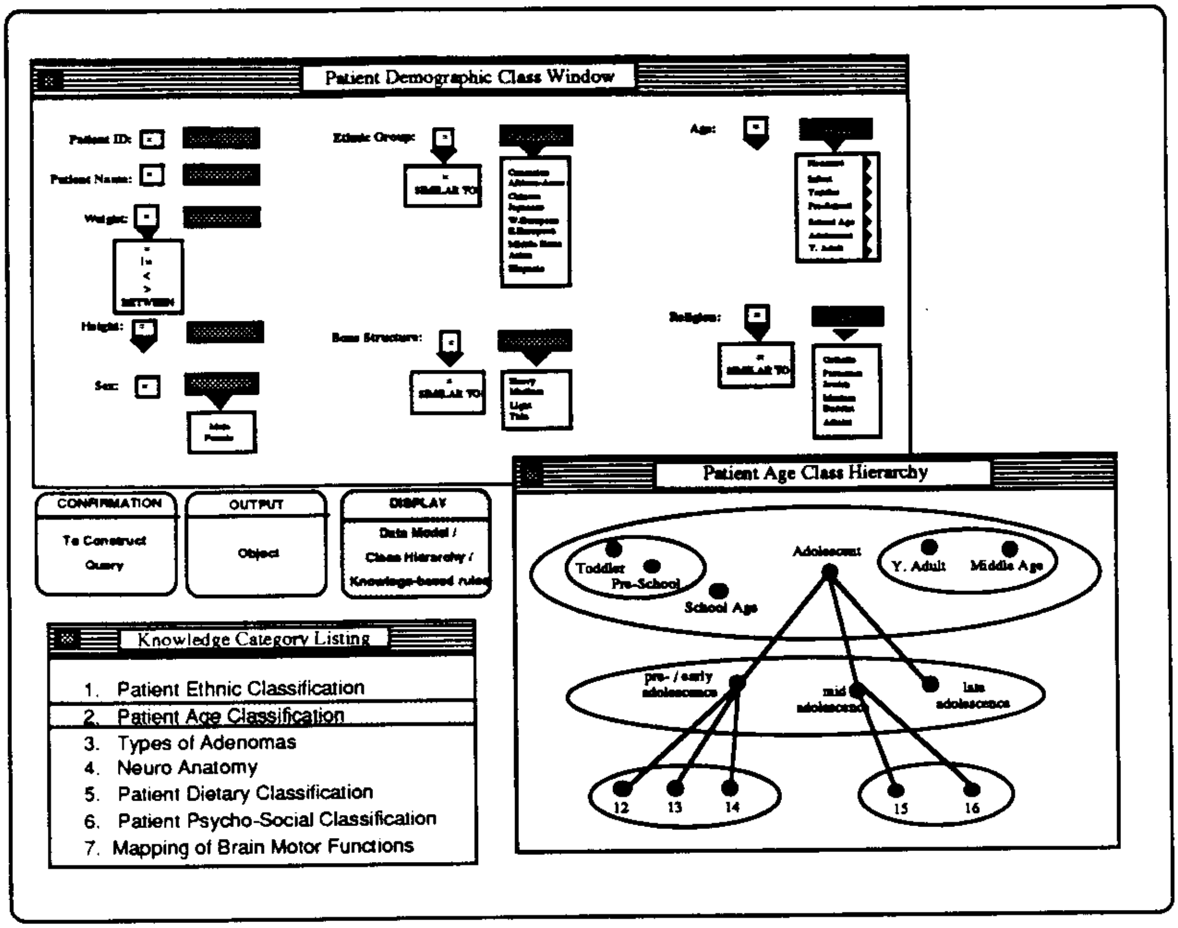


Fig. 3 - *Menu PICQUERY+ user interface. Note the user's ability to visualize knowledge including various type abstraction hierarchies.*

methods, and visualization methods for results. PICQUERY+ supports the User Requirements specified above. Figure 3 also illustrates the ability to visualize knowledge including various type abstraction hierarchies as the user defines the query.

Using PICQUERY+ we are able to conveniently express, for example, the following high level medical imaging queries:

- *Show brain cases which demonstrate abnormally shaped ventricles (see Fig. 1).*
- *Find an image of the proximal phalanx of the fifth finger for patient A.J. Smith and obtain the length of the major axis of this bone.*
- *Retrieve objects in image P2 similar in shape to brain objects considered abnormal in image P1.*
- *Show in a movie loop all images from patient cases which demonstrate the fusion phases of the metacarpal thumb ephiphysis and tubular bones.*

IMPRECISE AND CONCEPTUAL QUERIES

Traditional query processing methods accept precisely specified queries and provide only exact answers, thus requiring users to understand the problem domain and database schema. They return limited or null information if exact answers are not available. We have developed a **cooperative query answering** layer providing approximate, summary and conceptual answers to the original query (within a certain semantic distance to the exact answer) [Chu91]. Furthermore, information conceptually related to but not explicitly asked by the query is provided as part of the query solution.

We use a structured approach representing domain knowledge in type abstraction (concept) hierarchies. We then derive approximate answers using generalization and specialization relations from within a given concept hierarchy and association relations between different hierarchies [Chu92b].

For example, consider the query: *"Find the best treatment method for a tumor of type x for 12 year-old Korean boys"*. If no answer is available, then an approximate is found by relaxing various key concepts including tumor type, patient age, patient sex, and patient race. A more general query can be formulated which relaxes the tumor type and patient race. The modified query becomes: *"Find the best treatment methods for tumors in class X (of which little x is a member), on preteen Asians"*. Further associative information can also be provided such as the *success rate, side effects and cost of treatment* for the returned treatment protocol.

We have constructed the type abstraction hierarchies for several medical subject areas including the bone structure of the hand, disease classes of the hand, ethnic groups, age groups (see Figure 3), etc.

KNOWLEDGE DISCOVERY

Domain and expert knowledge are required for providing approximate, summary and conceptual query answers. New innovative clustering techniques based on machine learning methods have been developed to cluster similar objects measured by semantic distances, as well as objects that have near-arithmetic feature values (e.g., classification of the length of a hand bone for a certain age and sex group into long and short bones) [Merzb92]. These clustering techniques allow us to automatically construct type abstraction hierarchies for cooperative query answering for medical databases.

Work has begun on constructing type hierarchies for information stored in medical image databases. For example, with the above techniques we can derive: maturity differences and similarities in patients of a given sex and ethnic group with respect to growth rates of certain hand bones; the range of the length of a specified hand bone considered to be "thin" or "stubby"; the relationship of a patient's family history and his/her skeletal growth patterns.

IMPLEMENTATION

An object-oriented development environment is used due to the complexity of objects to be modeled, query requirements, and the specialization of various image processing methods. We use Gemstone (Servio Corp.) on Unix as the object-oriented DBMS, with a gateway to the UCLA PACS DBMS which uses Sybase. The programming environment uses Object/Visual Works (ParcPlace Systems). The database management programming language is OPAL (a database derivative of SmallTalk). Graphical data modeling is supported by Design/IDEF (Meta Software Corp.). The C language is used for image processing and visualization methods. Our heterogeneous distributed DBMS architecture, including the cooperative query answering layer, is the underlying multi-level architecture of KMeD.

TRANSFER OF TECHNOLOGY

Our software is modular and includes portability considerations in the design. It is therefore portable to any SUN Sparc workstation running the Gemstone database server. A user's guide explains our user interface PICQUERY$^+$ and contains details of the data models for the hand and brain applications. This document will be available for distribution to other interested researchers.

REFERENCES

[Carde93] A.F. Cardenas, I.T. Ieong, R.K. Taira, R. Barker, C.M. Breant, "The Knowledge-Based Object-Oriented PICQUERY$^+$ Language", *IEEE Transactions on Knowledge and Data Engineering*, August 1993.

[Chu91] W.W. Chu, Q. Chen, and R. Lee, "Cooperative Query Answering via Type Abstraction Hierarchy", in Cooperating Knowledge Based Systems, 1990, S.M. Deen (eds), Elsevier Science Publishing Co., Inc., 1991.

[Chu92a] W.W. Chu, I.T. Ieong, R.K. Taira, C.M. Breant, "A Temporal Evolutionary Object-Oriented Data Model and Its Query Language for Medical Image Management", *Proceedings of the Very Large Database (VLDB) Conference*, Vancouver, Canada, 1992.

[Chu92b] W.W. Chu and Q. Chen, "Neighborhood and Associative Query Answering", *J of Intelligent Information Systems: Integrating Artificial Intelligence and Database Technologies*, Dec. 1992.

[Josep88] T. Joseph and A.F. Cardenas, "PICQUERY: A High Level Query Language for Pictorial Database Management", *IEEE Transactions on Software Engineering*, vol. 14, no. 5, pp. 630-638, 1988.

[Merzb92] M. Merzbacher and W.W. Chu, "Instance-Based Clustering for Databases", *Proceedings of the 3rd ASIS Workshop on Classification Research*, Pittsburgh, PA, October 1992.

[Pietk91] E. Pietka, M.F. McNitt-Gray, M.L. Kuo, H.K. Huang, "Computer Assisted Phalangeal Analysis in Skeletal Age Assessment", *IEEE Trans.on Medical Imaging*, TMI-121-1991.