Organization and Retrieval in a Pictorial Digital Library

Yuri Quintana
New Media Lab
Graduate School of Library and Information Science
University of Western Ontario
London, Ontario, Canada
yquint@julian.uwo.ca
http://newmedia.slis.uwo.ca/

Abstract

This paper describes a knowledge based approach to organizing and retrieving pictures. Methods are described for using the human perceptions of pictures to create a frame knowledge base that represents the semantic content of pictures. A pictorial knowledge base management system is described that uses a conceptual clustering algorithm to learn new conceptual categories of pictures. A comparison is done between machine and human created conceptual categories. The paper shows how a frame knowledge based system can be used for intelligent picture retrieval using learned categories.

1 Introduction

A picture is a photograph, painting, or diagram that depicts individuals and objects in a particular situation. Organizing and retrieving pictures poses particular problems since each person will have their own interpretation of the meaning of a picture or a group of pictures. Most current picture retrieval techniques use keyword descriptions or syntactic characteristics such as colour and textures. There have been relatively few attempts at representing the semantic content of pictures. This knowledge representation could be used to retrieve pictures based on the semantic relationships between the observable entities in a picture. The results of a search could be displayed with bit-maps.

This paper describes the architecture of a knowledge-based management system that can be used to create and organize picture description knowledge bases. A conceptual clustering algorithm is used to organize the knowledge base into hierarchical classes of semantically related pictures.

DL 97 Philadelphia PA, USA
Copyright 1997 ACM 0-89791-868-1/97/7..$3.50

The paper shows how pictures can be represented using frames based on human perceptions, and how pictures can be retrieved from an organized picture knowledge base. The system was tested using pictures from news magazines.

This paper is organized as follows. Section 2 describes some current approaches to picture organization and retrieval. Section 3 describes the a picture knowledge base management system. Section 4 describes the design of a picture knowledge base. Section 5 describes how the knowledge base can be conceptually organized. Section 6 describes picture retrieval techniques from an organized knowledge base. Section 7 presents conclusions from this research and recommendations for future work.

2 Current Pictorial Information Systems

Current pictorial information systems have focused on methods for representing the shape, size, and orientation of objects in diagrams, as well as spatial knowledge. Retrieval of pictures has been primarily based on keyword searches of text descriptions or graphical queries on relational databases.

One of the simplest ways to retrieve pictures is with a keyword search on a database that contains text descriptions of pictures. For example, Pau [25] describes a system called PICTURE-BASIS that allows users to perform keyword searches. Users can then sort the results based on multiple fields, in ascending or descending order, and can create a new data set with the result of a sort. Al-Hawamdeh, Suliman, and Ooi [1] describe a menu-based natural language interface that allows users to pose boolean search queries on text descriptions of pictures.

Other pictorial information systems represent spatial relations of objects. Some of these systems analyze images syntactically, and then use spatial knowledge constraints to determine spatial relationships among objects. For example, Rabitti and Savino [27] describe a system for converting pixel images of floor plans into symbolic representations. An edge-detection algorithm is used to do a low level analysis of the image. A rule-based reasoning system is then used to recognize objects such as a chair and a table, as well as complex objects such as a dining room.

Hussmann and Schefe [14] describe a natural language interface to a knowledge base about floor plan diagrams. The diagrams consist of black and white lines on a white background depicting chairs and tables. Users can pose questions in German about the contents of the room such as "*Ist der Tisch klein (Is the table small ?)*". The system is capable of maintaining a dialog with the user. An example response in such a dialog is: "*Welchen Tisch Meinst Du ? (Which table do you mean)*". Inference rules are used to reason about spatial knowledge in this domain.

Other systems focus mainly on the structural representation of objects. For example, Cassi et. al. [5] describe the use of hypergraphs to represent the structure of constellations in space photos. A hypergraph is a representation which consists of graphs (nodes connected with labelled vertices), and hyperarcs which are vertices that connect two or more graphs. A relational database is used to store the graphs and picture retrieval is based on relational database queries.

Some pictorial systems allow database designers to define spatial constraints using graphical interface techniques. For example, Pizano et. al. [26] describe PICSL (Picture Integrity Constraint Specification Language), a language for graphically specifying spatial integrity constraints in pictorial databases. Integrity constraints are assertions that dictate the correct behaviour of databases. An example of a spatial constraint in a traffic database might be "state highways must not cross state boundaries". Users can graphically select icons to draw the spatial integrity constraints. These graphical diagrams are then converted into first order calculus expressions which are represented in an extended entity-relationship model. The constraints can then be used to maintain database integrity.

Another type of pictorial information system allows users to pose queries on aerial photos of land surfaces and geographical knowledge (also known as geographical information systems). Geographical knowledge can be stored in a relational database. For example, Lee and Fu [8] describe a system that allows queries about roads and cities to be entered graphically. The user can enter a query by drawing with a light pen device, and can select icons for items such as rivers, roads, and meadows. This query technique is known as *Query-by-Pictorial-Example* introduced by Chang and Fu in [6]. The system can answer questions like "*Find the portion of interstate that is enclosed within the boundaries of the city of Lafayette*". Queries are matched to a knowledge base about geographical features that was manually created and correspond to Landsat satellite images.

Kasturi et al. [16] describe another a system that stores geographical information about cities and roads. In this system, users can view pictures of maps and retrieve information by entering a natural language query such as "*Display all roads in the state which contains Fresno*". The parser translates the user's text into a database query of the form "*Entity , Contraint1, Constraint2, ...*". Each constraint specifies an attribute of an entity. Other similar systems can be found in [3; 10].

The research work has also been conducted on new data models and extensions to relational database query languages to handle pictorial data. For example, Joseph and Cardenas [15] describe the Pictorial Database Management System (PICDMS), and its query language PICQUERY, that were developed at UCLA. This system allows users to enter queries by filling in tables, an interaction style that is similar to *Query-By-Example* [6]. Pictorial data is represented in grid arrays, where the grid depth represents overlapping pictures. Variable length records are used to store all data values at one grid element.

Roussopoulos, Faloutsos, and Sellis [28] describe the PSQL query retrieval language for pictorial databases. PSQL is a pictorial extension of the SQL query language that can handle queries such as "show all cities with population > 100,00". The data is stored as abstract data types and relations can be defined over them. For example, in the relation State(state, population, capital, state-region), the first three are alphanumeric data types, and the last is a pictorial data type.

Leung, Hibler, and Mwara [20] describe a picture retrieval system that represents the content of a picture using a language called PDL (picture description language). PDL is based on an entity-attribute-relationship (EAR) data model, where entities, attributes, and relationships correspond respectively to the nouns, adjectives and verbs in the text descriptions of pictures.

3 A Picture Knowledge-Base Management System

A picture knowledge-based management system is described below that represents a deep and perceptual knowledge based on human elicited attributes. This includes knowledge about the observed entity's role, function, and activities. By adding more knowledge to the representation of pictures, intelligent query processing can be performed. Our system differs from previous work in the following ways: 1) it is a domain independent system that uses frames, a knowledge representation language, to represent the semantic content of pictures, 2) it can organize the pictures into conceptual categories and generate generalized descriptions of groups of pictures, and 3) it can perform guided searches for content-based picture queries in an organized knowledge base.

The term knowledge base management system (KBMS) is given to a system that stores, maintains, and organizes knowledge bases [4; 29]. The components of the picture knowledge base management system are shown in Figure 1. A knowledge base editor allows users to define primitive concepts in a semantic lexicon. These primitive concepts are then used to represent the semantic content of pictures in frame representations. The organization of pictures frames is done using an inductive generalization algorithm and a conceptual clustering algorithm. Clusters of conceptually related pictures are created. A knowledge base indexing scheme is used to link each picture in a cluster to its text description (caption), scanned image, and frame representation. An example of an organized picture knowledge base is shown in Figure 2.

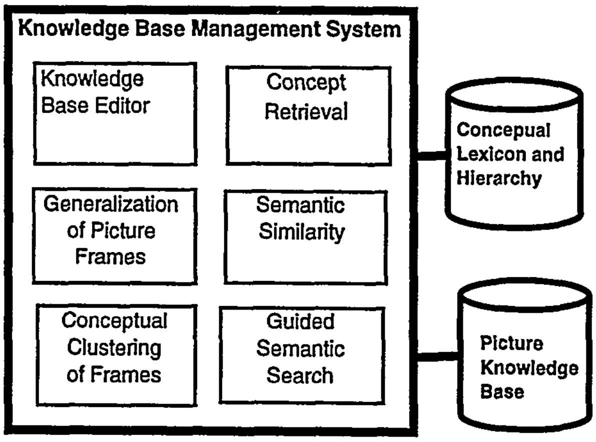

Figure 1: Picture Knowledge-Base Management System

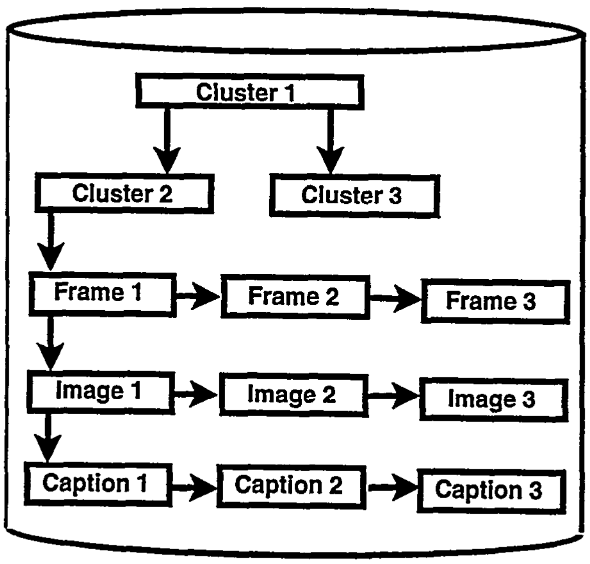

Figure 2: An Organized Picture Knowledge Base

4 Picture Knowledge Bases

A picture knowledge base was created to represent thirty-two photographs from Time-Life magazine. The pictures covered a wide variety of topics and themes. The knowledge base was used to store the descriptions of entities that are observable in a picture, and the relationships that exist among these entities.

In the design of such a knowledge base, it is necessary to determine a set of suitable symbols (primitive concepts) and a language for knowledge representation. There can be many entities in the foreground and background of a picture, and a knowledge engineer is capable of deciding what information is significant by using domain knowledge, common sense knowledge, and by taking into account the types of users and their anticipated queries. If pictures are to be represented in a knowledge base for retrieval by diverse users, there needs to be a diverse set of descriptions for each picture that is obtained from more than one person. A user's perception of a picture will influence they way he or she describes a picture to a picture retrieval system.

While some preliminary work has been done by other researchers in pixel-to-symbol conversion programs [27], it should be noted that the results to date are limited to specific domains. Thus, a domain-independent full scene analysis and conversion program to symbolic representation is still not feasible.

The knowledge engineering process described below is based on what humans remember about pictures. Although each picture can be individually captioned (a caption is the text below a picture), the caption represents only one person's perception and brief description of a picture. Three experiments were used to obtain further descriptions of the pictures; a short term memory and a long term memory recall experiment, and a picture organization experiment. In each case, humans were asked to describe the pictures in English for an unspecified future purpose. Humans were not given a specific purpose for the experiment so that the descriptions would not be biased towards one type of perception or description.

The short term memory recall experiment required human test subjects to look at each of the pictures. Immediately after all pictures had been viewed, the pictures were taken away and the test subjects were asked to recall as many pictures as they could. The test subjects were asked to describe as many details about each picture as they could remember. These written statements were considered to be an accurate reflection of the type of information that humans remember about pictures in short term memory.

Over a longer term (one day or more), humans were found to have very inaccurate descriptions of previously viewed pictures. Many test subjects recalled details in pictures which were not observable. Test subjects used reconstructive and episodic memory to recall pictures and describe details that they believe should have been in a picture, even if they weren't quite sure if these details were actually observed. Since we are interested in creating an accurate representation of a picture for retrieval purposes, the short-term memory recall descriptions were primarily used to build the knowledge base. Some details in the long term memory experiment were used only if they reflected accurate information about the picture. It is important to note however, that the retrieval system uses semantic similarity to retrieve pictures, and thus user queries which are not completely correct in their description of the intended picture may still retrieve the intended picture by semantic similarity, which is described in the picture retrieval section of our paper.

In the picture organization experiment, human test subjects[1] were asked to look at all the pictures and then group *related pictures* together. It was left up to the human test subjects to decide what was considered *related*

[1]These human test subjects were not the same group that were used in the short term memory recall experiment.

pictures. The grouping of pictures had to conform to the following constraints which were given in writing to each test subject:

- Each grouping must contain pictures that are related to each other.
- A picture can be in more than one grouping.
- There can be any number of groupings.
- Describe each group with an English sentence.

Test subjects described each group of pictures by using a list of numbers that identified the pictures in the group, and an English sentence that described the semantic content of the group.

After these three experiments had been conducted, each picture had several English descriptions which accounted for a diverse set of human perceptions (a captions, memory recall descriptions, and picture group descriptions). Some of these descriptions contained more information than what was immediately observable in the picture since humans added information that they previously knew about the observed event or entity (background knowledge). An example is recognizing and individual's name, occupation, and the significance of the event. The English language descriptions from all experiments were then used to create a set of concepts; symbols that have a unique semantic meaning. These concepts are the basis for the unambiguous representation of pictures.

It was found that as more pictures were added to the knowledge base, additional concepts were introduced. In order to consistently represent all pictures with the same level detail, any changes, such as the definition of new concepts, required that these changes be applied to previously represented pictures. This was not always done by the knowledge engineers, and this situation sometimes lead to inconsistent levels of details in the knowledge base. Ideally, this knowledge engineering methodology should be regressive each time new concepts or slots are added.

A frame representation was chosen to represent pictures because of their semantic expressiveness and their suitability for representation of pictorial information [9; 13; 23]. A frame is a data structure that can include both declarative and procedural information in pre-defined relations [2]. A frame is comprised of slots and fillers. Slots are used to describe some aspect of an object or entity. Slots can appear in any order in the frame, and more than once, except for the *frame_id* slot which must be the first slot in the frame (this requirement is particular to our frame representation). Slots are comprised of a slot name, and an ordered list of fillers. The requirement that fillers appear in a given order in a slot is also a specific requirement of our representation.

Fillers are concepts (symbols) that further define and constraint a slot. The order of fillers in a slot is important since fillers represent an ordered relation. For example, in the *clothing* slot, the first filler in the slot is a concept of a person and the second filler is a concept of some type of clothing that is worn by the person (and not vice versa). Figure 3 show examples of picture frames. Individual names of entities begin with a capital letter, and all other concepts begin with a lowercase letter.

```
frame:  Panama_riots
comment: riot in the streets of Panama
instance_of: riot
isa:  X police
set: X card 10
isa: Y protester
physical_location:  city_street
geographical_location: Panama
isa: T tree
set: T card 10
isa: C car
set: C card 4
hact: X marching
hact: Y running

frame: riot
comment: riots have protesters in street
comment: more protesters than police
instance_of: conflict_event
isa: Y protester
set: Y card NY
isa: X police
set: X card NX
value: NY > NX
physical_location: city_street
```

Figure 3: Frame for a picture that shows riots in Panama

Frames can be organized hierarchically such that one frame can inherit the properties of another. For example, a prototypical frame about riots would describe typical features of riots. Frames about specific riots would be instances of the prototypical frame for riots. Thus, common properties about riots would be represented only once in the prototype frame and instance frames would inherit these properties. Figure 3 shows a frame of a picture that shows riots in the streets of Panama City. The slot *instance_of* defines that this frame is an instance of the prototype frame for riot.

The concepts (fillers) used in the frames are stored in a concept hierarchy for use during picture retrieval. An example of this is specifying that the concept gavel is a specialization of the concept hammer (a gavel is a wooden hammer used by a judge in a courtroom). When searching for a picture that shows a person with hammer, our retrieval system is able to find a picture that shows a man with gavel (i.e. a specialized instance in the knowledge base satisfies a more general query). By storing the concepts in a concept hierarchy, retrieval is enhanced, as will be further illustrated in later sections.

The knowledge engineering methodology presented in the previous section shows how a set of concepts can be determined for representing pictures. Once a set of concepts were defined, a set of slot definitions were created. Slot definitions define how concepts can be ordered in a frame.

There are several inconsistencies that may be introduced during the creation of a picture knowledge base.

The first has already been mentioned and it deals with using a consistent level of detail to represent pictures. For example, consider the addition of a new picture which requires the definition of a new slot regarding observable weather conditions. Since this slot has just been defined, then it has not been used in any previously stored picture representations. However, a user may later pose a query using this slot believing that the slot has been used in all frames where there are observable weather conditions. Thus, it is important to revise the frame representations when new slots are defined.

It was found that the knowledge engineers did not always apply new slots to previously stored picture descriptions. This may have been due to lack of desire to review all previous frames, but also due to the fact that the frames in the knowledge base do not have a recorded time that indicates the time of last modification or revision. While it may be ideal to define all slots initially, our experience was that as new pictures and knowledge were added to the knowledge base, often it revealed new ways to represent details not previously deemed useful. Dynamic knowledge bases thus require dynamic and evolving representations. An improvement to our current knowledge-base management system would be to record the time when each frame was revised or added.

Another problem deals with the definition of new concepts. Concepts are used as the fillers in slots. In our system, all concepts are organized into a concept hierarchy which is used during picture retrieval. Although the concept hierarchy improves retrieval performance, maintaining the concept hierarchy is difficult. The addition of a new concept requires that it is classified in the concept hierarchy. As the concept hierarchy grows, the classification of new concepts becomes more difficult since the knowledge engineer may not necessarily be familiar with the entire structure of the concept hierarchy. The semantic similarity function described in the later sections of our paper could possibly be used to classify new concepts to the existing concept hierarchy.

Another problem deals with the representational consistency of queries and stored frames. Picture retrieval is based on the matching of query frames to picture frames in the knowledge base. Thus, it is necessary that query frames be represented using the same set of concepts that were used to represent the pictures in the knowledge base. In our experience with building the picture knowledge base, new concepts and frame definitions were constantly being added, renamed, deleted or modified. When a concept is modified or deleted, any frame that used that concept needs to be revised. Hence, consistency between query and picture representation becomes an issue.

5 Conceptual Organization of Pictures

The conceptual clustering algorithm organizes pictures not only on the observable features, but also on implicit properties. The algorithm uses inheritance reasoning to take into account background knowledge in the clustering. Prior to clustering, properties of concepts in the slots of picture frames are inherited from the background knowledge base. The algorithm creates clusters based on a group similarity function which maximizes the semantic similarity of all members of a cluster to each other, as opposed to maximizing the similarity of members to a randomly selected seed. This approach to clustering is based on a psychology theory of picture perception known as *gestalt.*

Picture frames are represented with concepts that are stored in a concept hierarchy, *CH,* which is a source of background knowledge. A background knowledge base, *BKB,* is used to add default and prototypical knowledge to each frame prior to making generalizations. Thus, clustering based not only on the explicitly stated features, but also on features that can be inherited or assumed by default.

A generalization algorithm is used to find common features in knowledge constructs, where a knowledge construct is the smallest unit of knowledge to be organized. In this case, the knowledge constructs are frames. However other types knowledge constructs can also be used (such as rules). The generalization algorithm uses a best first search to look for common features in knowledge constructs. The best generalizations are selected using a semantic similarity function, defined in Figure 4. This function uses a weighted combination of four semantic similarity heuristics: the common concepts score (CC),the generalization score (GS), a type distance score (TD), and a frequency of concepts score (FQ). The CC heuristic calculates a similarity value based on the number of concepts occurring in two knowledge constructs. The GS heuristic calculates a similarity value based on the size of the generalization. The size of a generalization, in the case of frame representations, is the number of generalized slots. The TD heuristic calculates a similarity value based on the average type distance of generalized concepts. The type distance is the number of links in the concept hierarchy from two concepts c_i and c_j to a common parent c_k. The FQ heuristic calculates a similarity value based on the frequency of concepts in the input knowledge base. The scores are then combined with a weigthed sum, as shown in Figure 4 such the result is a semantic similarity number from 0 to 1.

$$SS=\frac{WCCxCC + WGSxGS + WTDxTD + WFQxFQ}{WCC + WGS + WTD + WFQ}$$

Figure 4: The Semantic Similarity Function

Pairwise semantic similarity computations are done on all knowledge constructs to be clustered and stored in a table for later retrieval. The algorithm then begins to create conceptual clusters. Beginning with the highest semantic similarity found between two pairs of knowledge constructs, a cluster is created whose members are KC_i and KC_j, and whose conceptual description is their generalization KC_g. The process continues with decreasing semantic similarities found in generalizations. For each generalization, the knowledge constructs that created the generalization are either added to a previous cluster, or are used to create a new cluster.

When two knowledge constructs, whose generalization is K_g, are considered for inclusion in an existing cluster, a function called Group_Semantic_Similarity, defined in Figure 5, is used to compute the semantic similarity between all members of a cluster. The Group_Semantic_Similarity function calls the Semantic_Similarity_Lookup function to retrieve previously computed semantic similarity scores between any two knowledge constructs. This group semantic similarity is a measure of the *conceptual cohesiveness* of a group, that is to say how closely the members of a cluster relate to each other. This is a *gesltalt*-based similarity function, since it is more important how the members of the cluster relate to each other, than how any two items relate to each other. The theory of "*gestalt*" [17; 18] has its origins in the german word "*gestelt*", which refers to the way things have been "*placed*" or "*arranged*". This group similarity function biases the creation of new clusters toward what is maximally common about the group.

```
For all KC_i and KC_j are members of cluster KC_new {
   if( KC_i ≠ KC_j ) {
      Group_ss = Group_ss +
         Semantic_Similarity_LookUp( KC_i , KC_j )
      num_cluster_members =
         num_cluster_members + 1
      }
   }
return( Group_ss / num_cluster_members )
}
```

Figure 5: Definition of Group Semantic Similarity

The cluster descriptions created by humans and the conceptual clustering algorithm are shown in Figure 6 and 7 respectively. Figure 6 shows the number of humans that created a category of pictures. Figure 7 shows the number of pictures in each category created by the conceptual clustering algorithm.

Frequency	Cluster Descriptions
10	Sports
8	Politics
4	Children
3	Entertainment, Technology, Animal, Beauty Queens, Couples, Celebrations, Famous People
2	History, Family, People, Cinema, Celebration, War, Actors(es), Cartoons, Arts, Babies
1	Animals, Buildings, Cinema, Controversy, Culture, Dance, Demonstration, Events, Fashion, Flying, Future, Glamour, Groups, Lifestyle, Machines, Movies, Turning Points, Ugly, Wildlife, Zoology

Figure 6: Human Created Cluster Topics

The results of the conceptual clustering algorithm are quite comparable to humans. The conceptual clustering

Cluster Description	# of Pictures
American Married Couples	9
Famous Black American Men	7
California Scenes	5
Smiling People	5
Body of Water	4
War Scene and Casualty	3

Figure 7: Machine Generated Cluster Topics

algorithm created 6 clusters versus 8.2 clusters created by humans. Each machine-created cluster had on average 5.5 members, versus 5.3 in human created clusters. The machine-created clusters had a level of disjointness of 75% versus 60% for humans. Disjointness refers to the classification of pictures into more than one category. Of the six machine created clusters, 4 were also found by humans, namely *War Scene* , *Couples, Fashion*, and *Famous People.* The machine created two categories not found by the ten humans, namely *California Scenes*, and *Smiling People.* The latter is close to the human created category *People.*

A *prototypical* similarity matrix, P, was created based on the clustering data of humans, where each cell entry contains a 1 if more than half the test subjects had the two pictures in the same cluster. A similarity matrix, M, was created for the clustering solution created by the conceptual clustering program. The absolute difference error between a human prototypical similarity matrix, P, and a machine matrix, M, was found to be 19%. When comparing this value to the error values of each human test subject's clusters, the error is close to the human performance. The average human error was 22% (Std.Dev.=11).

In order to evaluate the degree of overlap between human and machine clusters, the *cluster set difference* function was used (Figure 8). A cluster set difference of 0 indicates that two sets of clusters are identical, and 100% indicates that two sets of clusters are completely dissimilar. The average cluster set difference between the machine and the human clusters was 38%. This is comparable to the cluster set difference of each human test subject when comparing it to other human created clusters. The average cluster set difference of the human test subjects was 29% (Std.Dev.=5).

6 Picture Retrieval

In this section, the conceptual clustering algorithm is evaluated in terms of improvements in picture retrieval from the picture knowledge base. The clusters descriptions in the knowledge base are used for heuristic guided searches for processing queries. Clustering and retrieval experiments were conducted with and without background knowledge.

An experiment was conducted with 100 queries that were created from the verbal descriptions of pictures obtained from human test subjects who were asked to de-

Given: CA, CB, are clusters

$$CSDifference = \frac{1}{2}[\ csd(CA, CB) + csd(CB, CA)\ ]$$

$$csd(\ CA,\ CB) = \frac{1}{|\ CA\ |} \sum_{i=1}^{|CA|} min\_csd(\ CA_i,\ CB\ )$$

$$min\_csd(\ CA_i,\ CB\ ) = min(1 - \frac{|\ CA_i \cap CB_j\ |}{|\ CA_i\ |})$$

$$for\ all\ CB_j \in CB$$

Figure 8: Cluster Set Difference Function

scribe previously viewed pictures. The correct retrieval for each query was identified manually. In one experiment, the unorganized pictures knowledge base was used and it was searched sequentially. In the second experiment the picture knowledge base was conceptually clustered, and a guided search was done to retrieve pictures. The knowledge-based picture retrieval approach uses frames to represent queries which are matched to frames in the picture knowledge base.

In a sequential search of the unorganized picture knowledge base, a query is compared to all the picture frames in the knowledge base. When searching in the conceptually clustered knowledge base, a query is compared to the generalization frame of each cluster using the semantic similarity function (Figure 4). The most relevant clusters are selected, namely those where the semantic similarity of the query and the generalization frame exceed the semantic similarity threshold, S_T. This process is called *cluster matching*. Once a cluster has been selected, the query frame is compared to the frames in that cluster. If the members are generalization frames of clusters, then *cluster matching* is used to select the most relevant clusters. If the members are frames that represent pictures, then *frame matching* is used. The purpose of the clustering is to identify clusters of pictures that are relevant to the query and to search only in those clusters, thereby avoiding a search of the complete knowledge base.

For each of these approaches, the result of a search is a list of pictures that are considered relevant to the query. (Not all recalled pictures are necessarily relevant to the query). Each picture has a relevance score associated with it. The highest score identifies the picture that is considered most relevant to the query. A cut off was used to limit the number of pictures retrieved. Recall and Precision measures were used to evaluate the effectiveness of each picture retrieval method.

Case	CKB	Precision	Recall	Frame	Slot
1	NO	61%	100%	10,000	3,598,438
2	YES	74%	95%	3,342	1,183,196

The first case uses sequential searching in an unorganized knowledge base, and the second case uses guided searching in a conceptually clustered knowledge base. It was found that both the sequential and guided search techniques had high recall (100% and 95% respectively). Recall was less than 100% in the clustered knowledge base since occasionally there is case where a cluster member is sufficiently different from the cluster generalization frame such that a query intending to retrieve the member does not match the cluster generalization frame above the semantic similarity threshold, S_T. This is limitation occurs because the generalization frames retain only the most prominent features of its members.

The advantage of using a clustered knowledge base with a guided search is that precision is higher than sequential searching (74% versus 61%), and the computational speed is much faster than sequential searching. For example, the number of frame comparisons were about one third less when using a guided search than when using a sequential search. Both of these are the result of not having to search the entire knowledge base.

A cluster generalization frame created without frame expansion is sometimes not descriptive enough for some queries to be considered relevant to the cluster. An example is a query that request pictures that relate to casualties of war. If there are several pictures of war torn cities and homes but do not directly show casualties, then the query would not be considered relevant to the cluster generalization frame, and consequently the pictures in that cluster would not be retrieved. However, a picture that shows destroyed homes will typically imply that there were some casualties. This shows that without frame expansion, clustering can not be done such that there is high recall of pictures. The conclusion is that frame expansion with clustering yields the best combined performance of precision, recall and computational complexity of all the four cases.

7 Conclusions

This paper has shown how knowledge engineering techniques can be used to create picture knowledge bases. A conceptual clustering algorithm is used to enhance the conceptual categories for organizing picture knowledge bases. Guided search can then be used to retrieve pictures, reducing search time and increasing precision.

The primary drawback to this approach is the effort in creating knowledge bases, which is a general problem of knowledge-based systems. However, for highly consulted picture knowledge bases, the investment in knowledge engineering might be worth the added retrieval performance on an intelligent picture knowledge base system. Future work will develop re-usable knowledge bases and conceptual hierarchies that will facilitate the development of future knowledge bases.

References

[1] S. Al-Hawamdeh, and B.C. Ooi. 1991. "Semantic Based Query Formulation in a PAS", Conference on Intelligent Text and Image Handling (RIAO'91), Barcelona, Spain, April 2-5, 1991, pp. 917-933.

[2] Barr, Arron, and Edward A. Feigenbaum. 1981. The Handbook of Artificial Intelligence, Volume 1, Los Altos, California: William Kaufmann Inc.

[3] J. Brolio, B.A. Draper, J. R. Beveridge, and A.R. Hanson. 1989. "ISR: A Database for Symbolic Processing in Computer Vision", IEEE Computer, Vol. 22, No. 12, pp. 22-30.

[4] Brodie, Michael L. and John Mylopoulos, (Editors). 1986. On Knowledge Base Management Systems, Berlin, Germany: Springer-Verlag.

[5] I. Cassi, P. Mussio, M. Padula, M. Protti, and G. Tonolli. 1992. "Relational tools to manage pictorial information", Journal of the American Society for Information Science, Vol. 18, pp. 375-398.

[6] N.S. Chang, and K.S. Fu. 1980. "Query-by-Pictorial Example", IEEE Transactions on Software Engineering, Vol. SE-6, No. 6 (November), pp. 519-524.

[7] D. Fisher. 1987. "Knowledge Acquisition via Incremental Conceptual Clustering", Machine Learning, Vol. 2, No. 2, pp. 139-172.

[8] Lee, Y.C. and K.S. Fu. "Query Languages for Pictorial Database System", In: Natural Language Communication with Pictorial Information Systems, Leonard Bloc, (Editor). Springer-Verlag, Berlin, pp. 1-142.

[9] Fikes, Richard, and Tom Kehler. 1985. "The Role of Frame Based Representation in Representation and Reasoning", Communications of the ACM, Vol. 28, No. 9, pp. 904-920.

[10] Fischler, M.A., and R.C. Bolles. 1989. "Image Understanding Research at SRI International", Proceedings of DARPA Image Understanding Workshop, pp. 21-31.

[11] J.H. Gennari, P. Langley, and D. Fisher. 1989. "Models of Incremental Concept Formation", Artificial Intelligence, Vol. 40, No. 1-3., pp. 11-61.

[12] S.J. Hanson and M. Bauer. 1989. "Conceptual Clustering and Polymorphy", Machine Learning, Vol. 3, No. 4, pp. 343-372.

[13] P.J. Hayes. 1979. "The Logic for Frames", In: Frame Conceptions and Text Understanding. D. Metzing, (Editor), Walter de Gruyter, pp. 46-61.

[14] Hussmann, Michael, and Peter Schefe. 1984. "The Design of SWYSS, a Dialogue System for Scene Analysis", In: Natural Language Communication with Pictorial Information Systems, Leonard Bloc, (Editor), Springer-Verlag, Berlin, pp. 143-201.

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

[16] R. Kasturi, R. Fernandez, M.L. Amlani, and W. Feng. 1989. "Map Data Processing in Geographic Information Systems", IEEE Computer, Vol. 22, No. 12, pp. 10-21.

[17] Koffka, K. *Principles of Gestalt Psychology*, London, England: Routledge and Kegan Publishing Ltd, 1935.

[18] Kohler, W. 1929. *Gestalt Psychology*. New York, NY: Liveright Publishing Corp. 1929.

[19] M. Lebowitz. 1987. "Experiments with Incremental Concept Formation: UNIMEM", Machine Learning, Vol. 2, No. 2, pp. 103-138.

[20] C.H.C. Leung, D. Hibler, and N. Mwara. 1992. "Picture Retrieval by content description", Journal of the American Society for Information Science, Vol. 18, pp. 111-119.

[21] R.S. Michalski and R.E. Stepp. 1983. "Learning from Observations: Conceptual Clustering", In: Machine Learning an Artificial Intelligence Approach, R.S. Michalski, J.G. Carbonell, and T.M. Mitchell (Editors), Tioga, pp. 331-363.

[22] R.E. Stepp and R.S. Michalski. 1983. "Conceptual Clustering of Structured Objects: A Goal Oriented Approach", Artificial Intelligence, Vol. 28, No. 1, pp. 43-69.

[23] M. Minsky. 1975. "A framework for representing knowledge", In: The Psychology of Computer Vision, P. Winston (Editor), McGraw-Hill, pp. 211-277.

[24] W. Niblack, et. al., "The QBIC Project: Querying Images by Content Using Color, Texture and Shape", IBM Technical Report RJ-9203, IBM Almaden Research Center, 1993.

[25] L.F. Pau, 1986. "A Picture and Text Query and Archiving System", Pattern Recognition Letters, Vol. 4, pp. 477-480.

[26] A. Pizano, A. Klinger, and A. Cardenas. 1989. "Specification of Spatial Integrity Constraints in Pictorial Databases", IEEE Computer, Vol. 22, No. 12, pp. 59-71.

[27] F. Rabitti, and P. Savino. 1991. "Automatic Image Indexation and Retrieval", Conference on Intelligent Text and Image Handling (RIAO'91), Barcelona, Spain, April 2-5, 1991, pp. 864-884.

[28] N. Roussopoulos, C. Faloutsos, and T. Sellis. 1988. "An Efficient Pictorial Database System for PSQL", IEEE Transactions on Software Engineering,

[29] Schmidt, Joachim W. and Costantino Thanos, (Editors). 1989. In: Foundations of Knowledge Base Management. Berlin, Germany: Springer-Verlag.