SEMCOG: An Integration of SEMantics and COGnition-based Approaches for Image Retrieval

Wen-Syan Li[†] K. Selçuk Candan[‡*] Kyoji Hirata[†]

C&C Research Laboratories[†]
NEC USA, Inc.
110 Rio Robles, San Jose, CA 95134
Email: {wen,hirata}@ccrl.sj.nec.com

Department of Computer Science[‡]
University of Maryland
College Park, Maryland 20742
Email: candan@cs.umd.edu

Abstract

Image retrieval is a key issue for many image database applications. Existing approaches include browsing and keyword, semantics, and cognition-based query processing. We argue that image retrieval using either approach alone is not effective. We introduce a system, SEMCOG (SEMantics and COGnition-based image retrieval), that allows users to pose a query using both semantics expressions and image examples. In our system, a query, such as "Retrieve all images in which there is a man on the right of a building and the building looks like this sketch", can be posed. This paper presents the architecture of SEMCOG, CSQL (Cognition and Semantics-based Query Language), the query language used in SEMCOG, and query user interfaces. Example queries are used to illustrate the image retrieval process in SEMCOG.

Key words. Object-based image retrieval, Content-based image retrieval, image databases, semantics, contents, cognition, Query processing.

1 Introduction

Image retrieval is a key issue in image database applications. Although there are some systems that allow users to retrieve images based on semantics, such as described in [1], SQL-based languages provide support only for textual queries. Since image data is primarily visual and hard to describe precisely, it is increasingly important to support searches that are specified by visual examples.

We see the existing approaches to image retrieval as follows: (1) Browsing and navigation approach: In this approach, users are presented with directories where each directory contains a single category of images (e.g. business, people, and plants). The users can navigate through directories and select a particular one to browse through the images of the corresponding category. (2) Syntactic keyword-based approach: In this approach, users input keywords and, as a result, they are presented with the images whose captions or keyword lists contain specified keywords. (3) Descriptive semantics-based approach: In this approach, users specify semantics of the image they want to retrieve. This approach requires the semantics of the query and images to be explicit so that the query processor can determine the similarity. The scene-based image retrieval can be viewed as a type of semantics-based approach. (4) Cognition-based approach: This approach supports queries by content semantics. Users may pose queries by providing drawings or image examples. The system retrieves images which are similar to the example provided by the user.

1.1 Problem Statement

Since image data is semantically richer than the traditional data forms stored in databases and since it is visual, we think that the semantics-based and the cognition-based approaches are more suitable for image retrieval. We argue that most of the existing work, which are either semantics-based or cognition-based alone, are weak in terms of flexible query specification capabilities.

The semantics-based approach is good at image retrieval based on image semantics. However, since images are visual and hard to describe in detail using text, this approach has a weakness of a low visual expressive capability. Another weakness of the semantics-based approach is that the semantics of images must be specified explicitly in advance if the semantics can not be extracted automatically or efficiently.

The cognition-based approach, on the other hand, has an advantage of using visual examples. However, one disadvantage of using the cognition-based approach alone is its lower precision because users' drawings are usually not precise enough. Another weakness of this approach is that it can not support queries on generalized concepts. For example, if a user wants to retrieve images containing

*This work was performed when the author visited NEC, CCRL.

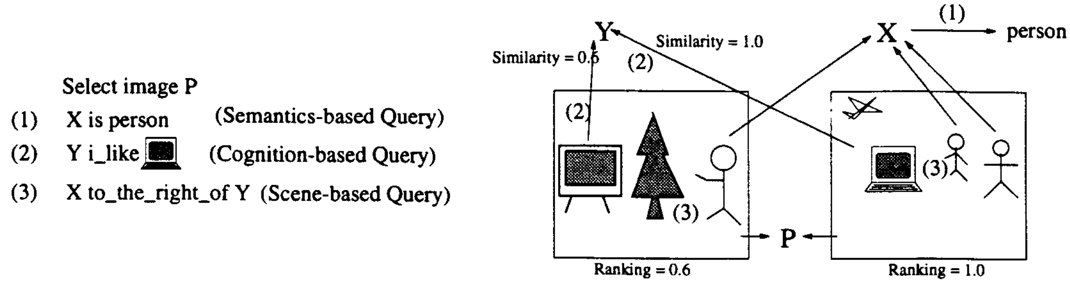


Figure 1: Example Query in SEMCOG

appliances, the user must provide drawings for all possible appliances, such as TVs, radios, computers, etc. This is not practical.

Another weakness of existing cognition-based image retrieval is that users cannot specify objects with finer granularity than a whole image. In many cases, users may want to retrieve an image based on semantics or visual characteristics of the objects contained in the image.

In this paper, we introduce a system, SEMCOG (SEMantics and COGnition-based image retrieval). SEMCOG aims at exploiting the advantages of both semantics and cognition-based approaches. SEMCOG allows users to pose a query "Retrieve all images in which there is a person who is to the right of an object that looks like this drawing". The query and its results are as shown in Figure 1. The arrows indicate how the variables X, Y, and P are bound with images, objects in images, and/or semantic meanings. In this example, three types of queries are involved: semantics-based (i.e. using textual specifications), cognition-based (i.e. using cognition specification), and scene-based (i.e. using object spatial relationship specification). SEMCOG is novel in allowing both semantics and cognition-based query specifications. As a result, the user query interface is more natural and flexible and the language expression capability is more powerful. Moreover, SEMCOG is also an object-based image retrieval approach, in which users can specify objects with granularity finer than a whole image.

The rest of this paper is organized as follows: We first give an overview the related work in this area. In Section 3, we present the architecture of SEMCOG and the components involved. In Section 4, we introduce the syntax of CSQL. In Section 5, we use two example queries to show the query processing and relaxation in our system. In Section 6, we report on the implementation status of the SEMCOG project. Finally we discuss the future work and offer our conclusions in Section 7.

2 Related Work

COIR (Content-Oriented Image Retrieval)[2] is a system for image retrieval developed at NEC, CCRL. The image retrieval technique used is based on features extracted from images, such as colors and shapes. Unlike most other existing approaches, where feature extraction and image matching techniques are based on the *whole* image, COIR's feature extraction and image matching techniques are object-based. This unique feature makes COIR more powerful and flexible since image matching can be applied to either whole images or particular objects in images.

Virage[3] is an image retrieval system based on visual features as image "primitives". These "primitives" can be general, such as color, shape, or texture, or domain specific. Virage transforms images from a data-rich representation of explicit image pixels to a more compact, semantic-rich representation of visually salient characteristics. These "primitives" are then used to solve specific image management problems.

QBIC (Query By Image Content)[4] is a system for image retrieval developed at the the IBM Almaden Research Center. It allows a user to query an image collection using features of image contents, such as colors, textures, shapes, locations, and layout of images. In QBIC, users pose queries by providing image examples.

Both Virage and QBIC support matching on whole images rather than objects in images like SEMCOG does. Furthermore, they do not support posing queries using combination of textual and cognitional specifications.

SCORE[5] is an on-going project for similarity-based image retrieval at University of Illinois at Chicago. This work focuses on using a refined ER model to represent the contents of pictures and the calculation of similarity values based on such representations. Users can use icons to specify objects, say *man* and *phone*, and then specify their relationships, say *to_the_right_of* and *holding*. The objects and their relationships are stored in tables. In their system, image retrieval is image modeling and query processing problems; image matching is not supported.

3 SEMCOG System Architecture

The architecture of SEMCOG is shown in Figure 2. SEMCOG contains five components: a *facilitator*, *COIR*, an *image semantics editor*, a *terminology manager*, and a *semantics-based query processor*. We now introduce the components' functionalities and the information maintained in SEMCOG.

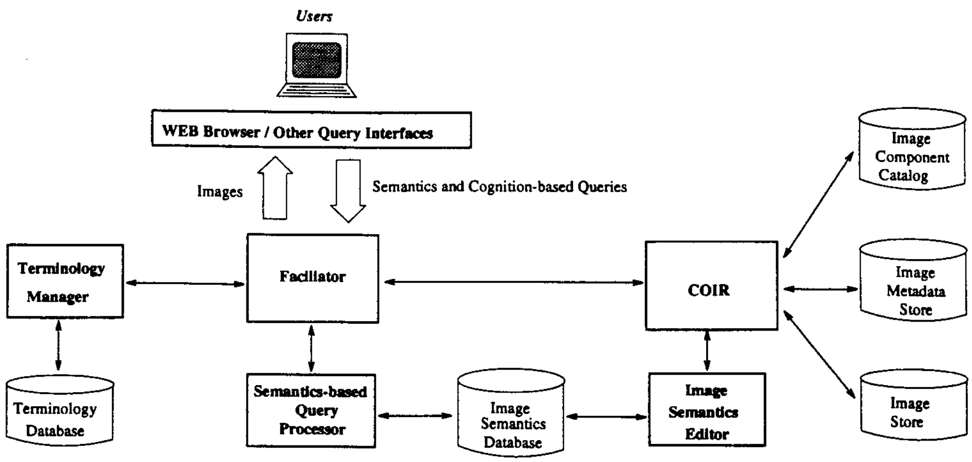


Figure 2: System architecture of SEMCOG

3.1 Components in SEMCOG

3.1.1 Facilitator

The *facilitator* coordinates the interactions between components of SEMCOG. It reformulates users' queries and forwards cognition-based query processing to COIR and non-cognition-based query processing to the *semantics-based query processor*. The query reformulation includes rewriting queries and changing sequences of predicates for query optimization purposes. We assigned these tasks to the *facilitator*, instead of the lower level components, such as the *semantics-based query processor*. One reason is that the *facilitator* has more complete knowledge of the query execution statistics and it can provide a globally optimum query processing. As a result, the *facilitator* can generate more efficient execution plan.

The query reformulation and relaxation requires consultation to the *terminology database* to resolve the terminology heterogeneity between users' terms and terms stored in the *image semantics database*. For example, a user may submit a query as "Retrieve all images in which there is an *appliance*." However, the term *appliance* is a generalized concept rather than an object. The *facilitator* consults the *terminology manager* to reformulate the query into "Retrieve all images in which there is a TV, a radio, or a computer."

3.1.2 COIR

COIR can retrieve images or image components with ranking based on similarity. COIR is an object-based approach. A region is defined as a homogeneous (in terms of colors) and continuous segment identified by COIR. A component (object) is an entity that is assigned a semantic meaning. A component may contain one or more regions. There can be one or more components in an image. We view a region as a physical object and a component as a logical object. The relationships between *image*, *object*, *component*, and *region* are illustrated using an example shown in Figure 3. Note that COIR, alone, can not identify the objects within an image. The main task of the COIR is to identify image regions. Since an object may consist of multiple image regions, COIR needs to consult an *image component catalog* for matching sets of image objects. COIR may "recommend" multiple corresponding objects because there are more than one matching objects in the *image component catalog*.

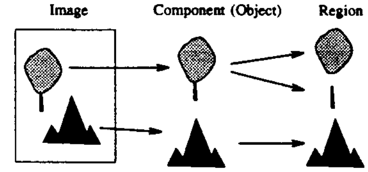


Figure 3: Image Composition in SEMCOG

3.1.3 Image semantics editor

The *image semantics editor* interacts with COIR to edit the semantics of an image. The interactions between COIR and the editor are shown in Figure 4. The steps involved in image semantics extraction and editing are as follows: (1) COIR identifies image regions of an image. (2) COIR recommends candidate components and their corresponding semantics by consulting the image component catalog. (3) The user confirms or modifies the results by COIR. (4) The user may select components to store in the image component catalog if he/she thinks these components can be used for "representative samples". The image component catalog is built incrementally in such way. (5) The semantics extracted is then stored in the image semantics database.

The image semantics database is conceptually visualized in Figure 5. The semantics of an image can be represented as a linked list of pointers pointing to the image and its components. Each image has a tag that contains a list of adjectives describing attributes associated with the whole image. An example of adjectives can be "elegant". "Elegant" can be associated with an image by examining the image's color combination. Each component also has tags describing its attributes, such as colors. COIR also

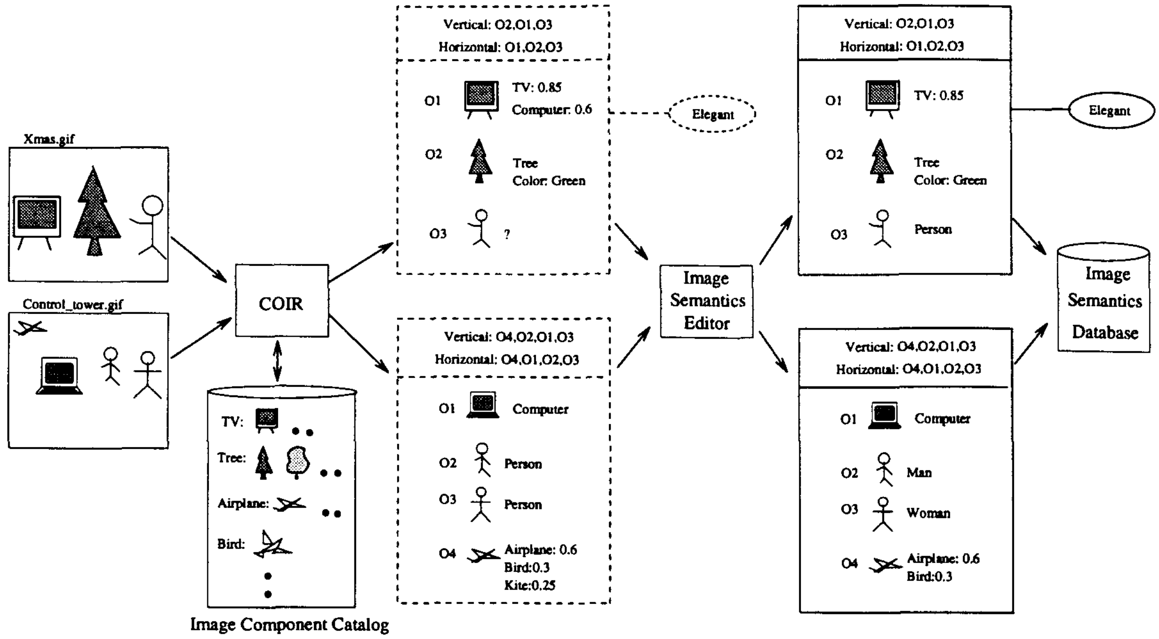


Figure 4: Image semantics extraction in SEMCOG

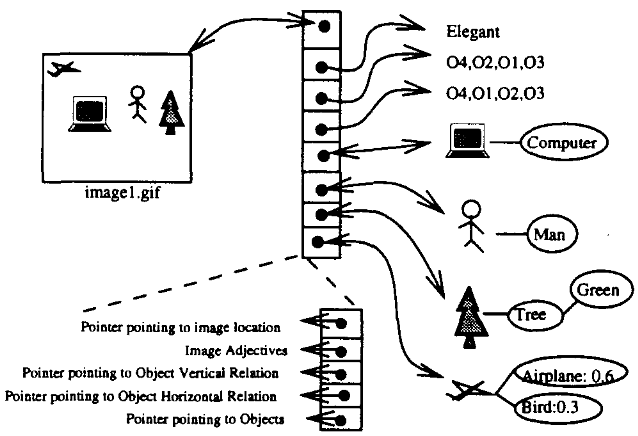


Figure 5: Image Semantics Database

identifies the vertical and horizontal relations between objects. These spatial relationships are identified based on the comparison of the four outermost points of objects' components.

Note that the image metadata store shown in Figure 2 stores pre-extracted image metadata, such as colors and shapes, and the *image component catalog* is built incrementally as *image semantics editor* users select a portion of image components and store them as "samples" in the *image component catalog.*

3.1.4 Terminology manager

The *terminology manager* maintains a terminology base that is used for query relaxation and term heterogeneity resolution. Given a term, the *terminology manager* can return multiple matching alternatives with ranking. These rankings are then used during the query processing in generating the final ranking of the results. Please note that we do not need to build a terminology database. *Webster* or more sophisticated databases, such as Wordnet[6], can be employed.

3.1.5 Semantics-based query processor

The *semantics-based query processor* provides answers to queries concerning the semantic information stored in images. The image semantics required for query processing is pre-extracted from the images using COIR and the *image semantics editor* and stored in the *image semantics database.*

3.2 Information maintained in SEMCOG

We summarize the information maintained in SEMCOG and its usage in Table 1. Our design aims at providing an open architecture that can be applied to other media, such as audio and video. The information maintained in SEMCOG can be categorized into media independent information, such as the *terminology database* and the image semantics database, and media dependent information, such as the *Image component catalog*, the *image metadata store*, and the *image store.* Our design has an advantage of seamlessly applying to other media by replacing the *Image component catalog*, the *image metadata store*, and the *image store* with components that are capable of handling other media.

Figure 5 visualizes the data structure of image semantics stored in SEMCOG. The information maintained in SEMCOG is actually stored as tables. For easy illustration purpose, we assume an image database *ImageDB*

Information	Maintained by	Used by	Purposes
Terminology database	Terminology manager	Facilitator	Query relaxation
Image semantics database	Image semantics editor	Semantics-based query processor	Semantics-based query processing
Image component catalog	COIR	COIR	Image component identification
Image metadata store	COIR	COIR	Cognition-based matching
Image store	COIR	COIR and facilitator	Image storage

Table 1: Information maintained in SEMCOG and Its Usage

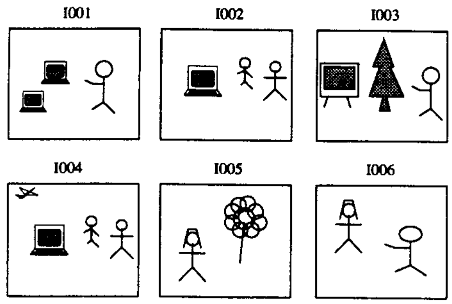


Figure 6: Images in *ImageDB*

Images	Adjectives	Objects			
I001		O101	O102	O103	
I002		O201	O202	O203	
I003	Elegant	O301	O302	O303	
I004		O401	O402	O403	O404
I005		O501	O502		
I006		O601	O602		

Table 2: Objects contained within the images

which consists of six images as shown in Figure 6. Table 2 shows the objects that are contained within each image. These objects are extracted from the images by the COIR system. The coordinates are pointers pointing to component in images. The properties of these objects are described in Table 3. Each object has a corresponding image, a set of coordinates describing its location within that image, and a set of semantics. These semantics are returned by the COIR system along with corresponding similarity values which range between 0.0 and 1.0. SEMCOG also maintains the spatial relationships between the objects in the database (as shown in Table 4).

We will use *ImageDB* later to illustrate the query processing in SEMCOG in Section 5.

4 Query Language

In this section, we give an overview CSQL (Cognition and Semantics-based Query Language). CSQL allows users to pose queries using the following criteria:

Semantics-based selection criteria: Users can specify semantic expressions to retrieve images. Given the semantic terminologies are not standardized, in many cases, users may not be able to provide precise semantics of the entities. Hence, in SEMCOG, we allow users to describe different levels of strictness for semantic relationships between entities using the *is*, *is_a*, and *s_like* binary predicates. The *is* predicate returns true if and only if both of the arguments have identical semantics (man vs. man, car vs. car). The *is_a* predicate returns true if and only if the second argument is a generalization of the first one (car vs. transportation, man vs. human). The *s_like* predicate returns true if and only if both arguments are semantically similar (man vs. woman, car vs. truck).

Object	Image	Semantics	Similarity	Semantics	Similarity
O101	I001	Computer	0.93	TV	0.65
O102	I001	Computer	0.82	Window	0.25
O103	I001	Man	0.75	Woman	0.69
O201	I002	Computer	0.98	TV	0.85
O202	I002	Man	0.88	Statue	0.45
O203	I002	Man	0.76	Statue	0.55
O301	I003	TV	0.85		
O302	I003	Tree	1.00		
O303	I003	Human	1.00		
O401	I004	Airplane	0.60	Bird	0.30
O402	I004	Computer	1.00		
O403	I004	Man	1.00		
O404	I004	Woman	1.00		
O501	I005	Woman	0.82	Statue	0.34
O502	I005	Flower	0.67	Bird	0.37
O601	I006	Woman	0.95	Child	0.66
O602	I006	Man	0.79	Chimpanzee	0.62

Table 3: Properties of the objects

Image	Orientation	Objects			
I001	horizontal	O101	O102	O103	
	vertical	O102	O103	O101	
I002	horizontal	O201	O202	O203	
	vertical	O201	O203	O202	
I003	horizontal	O301	O302	O303	
	vertical	O301	O302	O303	
I004	horizontal	O401	O402	O403	O404
	vertical	O401	O403	O402	O401
I005	horizontal	O501	O502		
	vertical	O502	O501		
I006	vertical	O601	O602		
	horizontal	O601	O602		

Table 4: Spatial relationships of the objects

Cognition-based selection criteria: Users can also specify a cognition based selection criteria to retrieve images. SEMCOG allows users to describe different cognition-based selection criteria between two entities using the *i_like* and *contains* binary predicates. For these two predicates, the system compares two entities based on their image identities. The first argument of the *contains* predicate must be a whole image. This predicate returns true if the second argument, which is an object, is contained in the first argument.

Scene-based selection criteria: As we introduced earlier in the section, CSQL also allows users to describe the spatial relationships between objects. For example, the *to_the_right_of* predicate takes two entities and checks whether the first object argument is to the right of the second object argument. Other scene-based predicates include *to_the_left_of*, *above_of*, and *below_of*. Please note that we only focus on these spatial relationships that can be *automatically* identified by COIR. Other spatial relationships, such as *in_front_of* and *behind*, are more difficult tasks for any image processing technique to perform fully automatically.

Besides the above selection criteria and predicates, CSQL also allows users to elaborate on their query specifications using *adjectives*: *color adjectives* and *image adjectives*. Color adjectives apply to entities, such as *red car*, and image adjectives apply to the whole image, such as *elegant* images. The *elegant* adjective has been implemented in COIR. An image is considered *elegant* based on the color combination of its regions. Please note that in this work, we only consider those adjectives that can be automatically identified by COIR, such as colors. Other adjectives such as large and small are more difficult to identify since these adjectives are *relative* relations between objects, not *absolute*.

SEMCOG allows its users to retrieve images based on combinations of the following criteria: image similarity, containment, and spatial relationship. This gives users a greater flexibility to pose image retrieval queries.

5 Example Queries

Now we show the query processing in SEMCOG using the example database in Section 3.2.

Example 5.1 *Retrieve all images in which there is a TV.* In CSQL syntax, this query can be posed as follows:

select image X *where* X *contains* TV

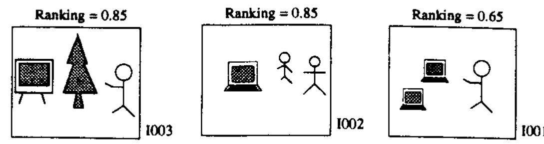
The processing of this query involves the following steps:

Step 1: Facilitator asks the semantic query processor to find all images that contain a TV.

Step 2: The semantics-based query processor checks the semantic information database to find all the images that contains a TV by checking with Table 3. The result of this step, IDs of the matching images, are returned to the facilitator with the corresponding similarity values as follows:

Object	Image	Semantics	Similarity	Semantics	Similarity
O101	I001	Computer	0.93	TV	0.65
O201	I002	Computer	0.98	TV	0.85
O301	I003	TV	0.85		

Step 3: The final results are images with their ranking as follows:



Example 5.2 *Retrieve all images in which there is a man and a transportation and this man is to the right of the transportation.* This query can be posed as follows:

select image X
where X *contains* Y
and X *contains* Z
and Y *is* man
and Z *is_a* transportation
and Y *to_the_right_of* Z *in* X

One possible execution plans is as follows: The facilitator

Step 1: creates one image entity for X (ix), two dual entities[1] for Y and Z (dy and dz respectively), and links these dual entities to the image entity ix.

Step 2: assigns *man* to the semantic component of dy.

Step 3: consults the semantics manager for a list transportations, and assigns *Airplane* and *Car* to dz.

Step 4: passes ix, dy, and dz to the semantics-based query processor for a list of triples that satisfies the *to_the_right_of... in* criterion. The query processor returns answers as long as both the spatial criterion and the links created in Step 1 are simultaneously satisfied.

Object	Image	Semantics	Similarity	Semantics	Similarity
O103	I001	Man	0.75	Woman	0.69
O202	I002	Man	0.88	Statue	0.45
O203	I002	Man	0.76	Statue	0.55
O403	I004	Man	1.00		
O602	I006	Man	0.79	Chimpanzee	0.62

Object	Image	Semantics	Similarity	Semantics	Similarity
O401	I004	Airplane	0.60	Bird	0.30

Here, $O403$ and $O401$ satisfy the link created in Step 1, i.e. both of them belong to the same image, $I004$. They also satisfy the spatial constraint:

Image	Orientation	Objects			
I004	horizontal	O401	O402	O403	O404

As a result of this step, ix gets bound to $I004$ with ranking of 0.6.

Step 5: returns the IDs of the images returned by the semantics-based query processor. Since the previous step returned only $I004$, the result of the query is as follows:

[1] A dual entity is an entity that has both semantic meaning and cognitional meaning.

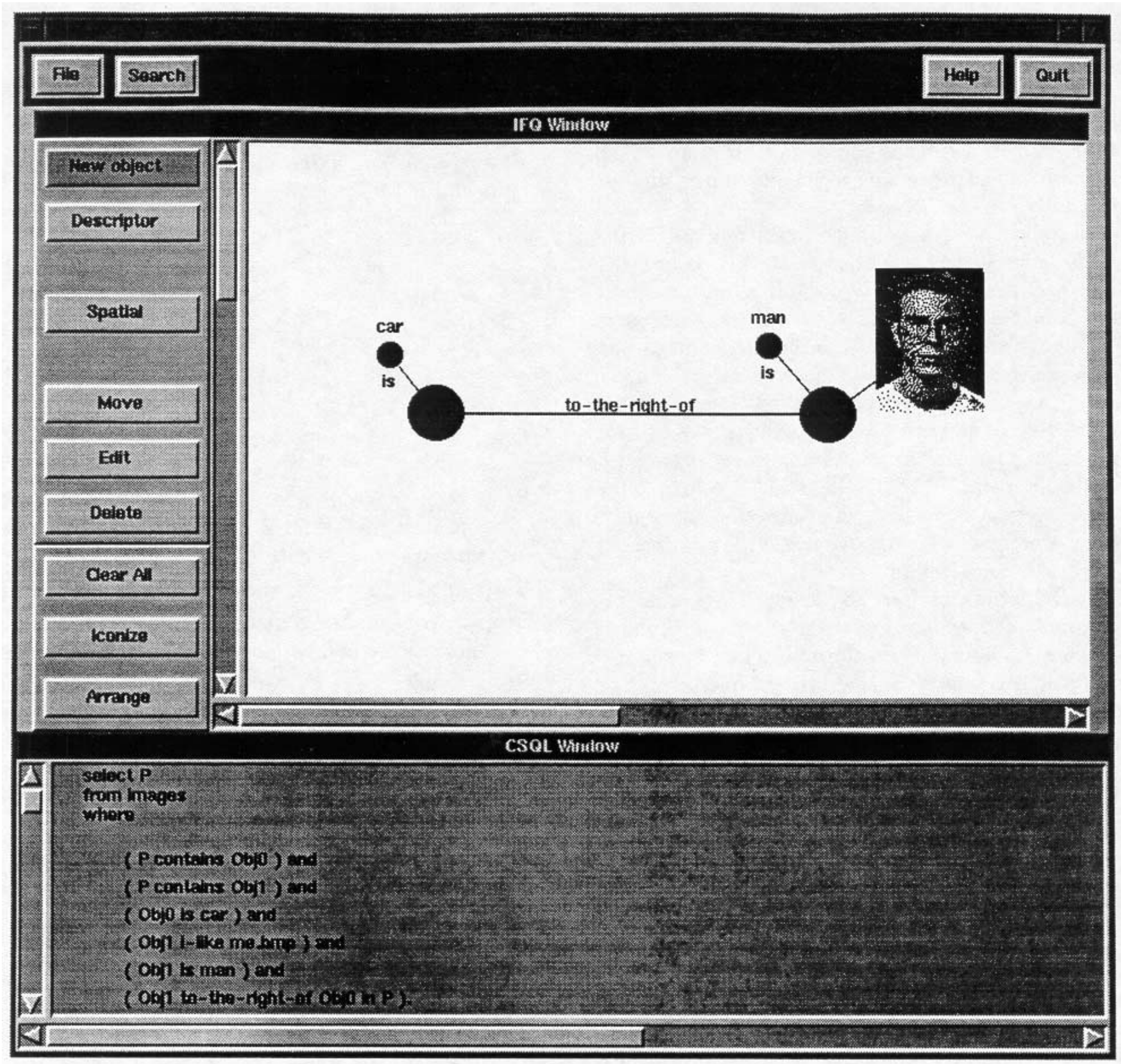
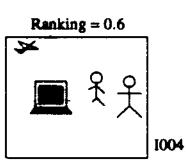


Figure 7: IFQ menu in SEMCOG



6 User Interfaces

SEMCOG currently supports a visually richer user query interface called IFQ (In Frame Query) that provides a more natural image query user interface. Figure 7 shows a query "Retrieve all images in which there is a man to the right of a car and this man looks like this image" using IFQ. In IFQ the user uses a bullet to represent an object and then associates additional *descriptors* to describe the entity. For example, the user uses "i_like < *image* >" to describe the entity *Man*. The user can also specify relations between entities. Queries using CSQL and IFQ are equivalent. As shown on the button of Figure 7, the CSQL corresponding to IFQ specified by the user is generated by SEMCOG automatically. Please note that the user needs not be aware of the syntax and variable specification in CSQL since these are handled by the IFQ interface.

7 Conclusions

SEMCOG is an on-going project at NEC, CCRL in San Jose. Currently we are building SEMCOG on object-relational database systems as a add-on module. In this paper, we present a novel approach to image retrieval by integrating semantics and and cognition-based approaches. We introduced SEMCOG (SEMantics and COGnition-based image retrieval) which provides a hybrid solution to the problem.

The contributions of our work include (1) Support both semantics and and cognition-based query specification. It is a more flexible and visually richer query language. (2) Define a language incorporating semantics, scene, and cognition-base query. (3) Our query language allows users to specify objects with granularity finer than a whole image. (4) SEMCOG provides an open architecture that can be easily extended to video and audio retrieval by extending COIR and the semantics editor with components which are capable of multimedia recognition. (5) IFQ provides a media-independent query interface for multimedia retrieval.

Acknowledgments

We would like to express our appreciation to referees for their thoughtful reviews and Yoshinori Hara for his invaluable comments and suggestions on this work.

Biborgaphy

Wen-Syan Li is currently a research staff member at Multimedia Software Department of Computers & Communications Research Laboratories, NEC USA Inc. He received his Ph.D. in Computer Science from Northwestern University in December 1995. His current research interests include multimedia data modeling, multimedia databases, image/video retrieval, digital libraries, heterogeneous database, and object-relational database systems.

K. Selçuk Candan is currently a Ph.D. candidate at the Department of Computer Science at the University of Maryland, College Park. He got his bachelors degree in 1993 from the Bilkent University, Turkey. His research interests are distributed multimedia authoring systems, video servers for video-on-demand systems, content based video indexing, query/retrieval of multimedia data, and query optimization issues in distributed heterogeneous databases.

Kyoji Hirata is currently a research staff member at Multimedia Software Department of Computers & Communications Research Laboratories, NEC USA, Inc. He received his B.E. degree in Electric Engineering and his M.E. degree in Electronic Engineering from the University of Tokyo in 1987 and 1989, respectively. He joined NEC Corporation in 1989, and has been engaged in the research and development of multimedia databases. His research interests are multimedia databases, content-based information retrieval, hypermedia systems and multimedia applications.

References

[1] W. W. Chu, T. Ieong, and R. Taira. A Semantic Modeling Approach for Image Retrieval by Contents. *VLDB Journal*, 1994.

[2] Kyoji Hirata, Yoshi Hara, H. Takano, and S. Kawasaki. Content-Oriented Integration in Hypermedia Systems. In *Proceedings of 1996 ACM Conference on Hypertext*, March 1996.

[3] J. R. Bach, C. Fuller, A. Gupta, A. Hampapur, B. Horowitz, R. Jain, and C.-F. Shu. The Virage Image Search Engine: An Open Framework for Image Management. In *Proceedings of the SPIE - The International Society for Optical Engineering: Storage and Retrieval for Still Image and Video Databases IV*, San Jose, CA, USA, February 1996.

[4] Jonathan Ashley, Myron Flickner, James Hafner, Wayne Niblack Denis Lee, and Dragutin Petkovic. The Query By Image Content (QBIC) System. In *Proceedings of the 1995 ACM SIGMOD Conference*, San Jose, California, May23-25 1995.

[5] Y. Alp Aslandogan, Chuck Thier Clement Yu, Chengwen Liu, and Krishnakumar R. Nair. Design, Implementation and Evaluation of SCORE. In *Proceedings of the 11th International Conference on Data Engineering*, Taipei, Taiwan, March 1995. IEEE.

[6] G. A. Miller. WordNet: A Lexical Databases for English. *Communications of the ACM*, pages 39–41, November 1995.