

IFQ: A Visual Query Interface for Object-based Image Retrieval

Wen-Syan Li K. Selçuk Candan Kyoji Hirata Yoshi Hara

C&C Research Laboratories, NEC USA, Inc., 110 Rio Robles, San Jose, CA 95134

Email: {wen,candan,hirata,hara}@ccrl.sj.nec.com

KEYWORDS: Image retrieval, visual query interface.

INTRODUCTION

There are two major directions for image retrieval interface development. One direction is manipulation through query languages, which is more precise to computers but not user friendly since users need to know database schema and query languages. Another disadvantage is that query languages do not visualize queries so that users can match the query representation with their mental models. Approaches include keyword-based interfaces and SQL-like query languages.

Another direction is to develop a more natural interface. Approaches include (1) cognition-based interfaces that support query by image examples, and (2) descriptive semantics-based approach in which users pose queries by describing target image semantics in a somewhat “natural language” fashion. This type of interfaces has an advantage of being natural to users, but it has disadvantage of low precision because queries are ambiguous to computers.

We argue that there is a gap between these two directions. We believe a query interface that can visualize queries and also support precise manipulation is essential to effective image retrieval. Moreover, image data is semantically richer than the traditional data forms. The fact that segments in an image have associated semantics and they constitute meaningful structures is ignored by most systems. Users should be able to query objects with finer granularity than a whole image.

The cognition-based approach can visualize queries but it can not support generalized concepts, such as transportation. The semantics-based approach is more appropriate in expressing generalized concepts. However it has a low visual expressive capability. We believe both visual and semantics expressive capabilities are essential in image retrieval.

IFQ APPROACH

Figure 1 shows an example in which a user wants to retrieve an image containing a person and a computer where the person is to the right of the computer. There is a gap between the concept in her mind and actual images stored in a database system. Her concept is more natural, but rather ambiguous to the database system. Effective retrieval of images requires a more precise query language. However, this type of manipulation language is usually not user friendly.

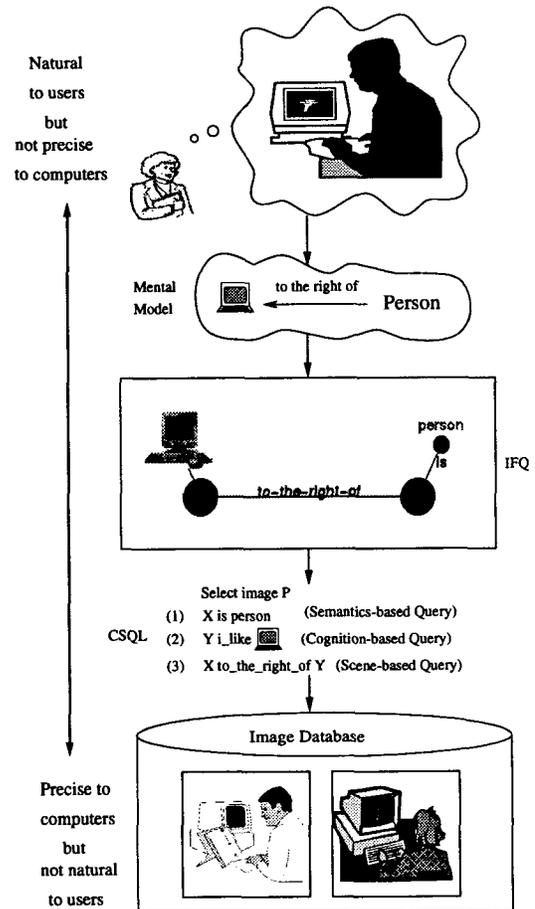


Figure 1: IFQ as a Visual Query Interface

IFQ (In Frame Query) is an attempt to bridge this gap. IFQ is a visual interface for SEMCOG (SEMantics and COGnition-based image retrieval system), an *object-based* image retrieval system. In SEMCOG, the queries are posed by specifying image objects and their layouts. IFQ allows users to query objects with *finer granularity than a whole image* and specify spatial relationships between objects. IFQ also aims at integrating semantics and cognition-based approaches to give users a higher flexibility to pose queries.

The user’s query can be simplified as a visualization model shown at the top of Figure 1. She then specifies this visualization model using IFQ using either visual examples or semantics. An actual window dump of IFQ for this query is shown in the middle of Figure 1. This query is then translated (by IFQ) to CSQL (Cognition and Semantics-based Query Language) query, a SQL-like query language used in SEMCOG. The CSQL query is precise to the database system.

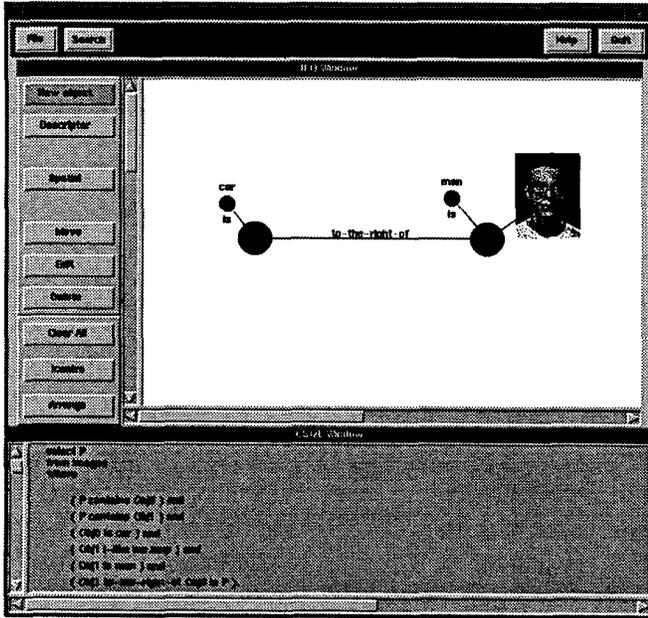


Figure 2: IFQ Visual Query Interface

IFQ QUERY INTERFACES

The image retrieval related predicates in CSQL include:

- *Semantics-based*: *is* (e.g. man vs. man), *is_a* (e.g. car vs. transportation, man vs. human), and *s_like* for "semantics like" (e.g. car vs. truck).
- *Cognition-based*: *i_like* for "image like" that compares visual signatures of two arguments and *contains*.
- *Scene-based*: *to_the_right_of* in, and etc.
- Image and object *adjectives*, such as *red car*.

In IFQ, objects are represented as bullets and *descriptors* are represented as small bullets, attached to these objects describe the properties of the corresponding objects. Figure 2 shows an IFQ query "Retrieve all images in which there is a man to the right of a car and he looks like this image". The IFQ query is posed as follows: The user introduces the first object in the image, and then further describes the object by attaching "i_like < image >" and "is man" descriptors. After a user specifies a image path or provide a drawing, the interface automatically replaces the descriptor with the thumbnail size image the user specifies. Then, the user introduces another object and describes it using the "is car" descriptor. Finally, the user describes the spatial relationship between these two objects by drawing a line, labeled by *to-the-right-of*, from the man object to the car object. Please note that while user is specifying the query using IFQ, the corresponding CSQL query is automatically generated in the CSQL window. Users simply pose queries by clicking buttons and dragging and dropping icons representing entities and descriptors.

Functionalities supported by IFQ include:

- **Checking and Arrangement**: IFQ can check the matching between the layout specifications provided by the user and the actual layout on the screen. If there is a mismatch, IFQ rearranges the query objects on the screen according to the query specifications. Furthermore, if there is an inconsistency in the layout specifications provided by the user (e.g. man is

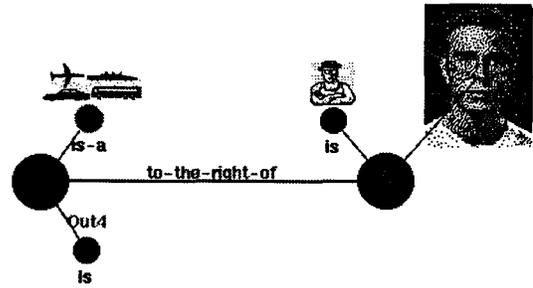


Figure 3: Relaxed Query for Extracting Semantics

Solution	Confidence	Image	Out4
1	0.760000		car
2	0.280000		bus

Figure 4: Query Result

above tree and tree is above man), IFQ informs the user by highlighting the inconsistency specifications.

- **Iconization**: *iconize* function can replace the semantics terms on the IFQ window with the corresponding icons. This improves the perceptual qualities of the query.
- **Query Parameter Specification**: Users can specify ratio of color importance to shape importance (used by COIR for image matching), ratio of object matching importance to structure matching importance (used to rank images), filtering parameters, such as a minimum degree of relevancy and a maximum number of image retrieved, and image related parameters, such as titles and authors.

Query results contain thumbnail size images ranked by their degrees of confidence. Users can click on any thumbnail image to see the real image as shown on the right side. However, in many cases, users do not have specific target images in mind. Users want to extract and browse semantic and information about image objects. This form of interactions is supported through extracting semantics and relaxing queries. Users can relax the condition of being a *car* to "being a kind of transportation", and introduce an unbound descriptor *out4* to check the actual semantics as shown in Figure 3. Figure 4 shows two candidate images including an image containing a bus. This image has a lower confidence, because the human in the image can not be identified as men.

This demonstration shows a novel interface for object-based image retrieval. IFQ integrates semantics and cognition-based retrieval approaches to give users a higher flexibility for describing objects in the query. IFQ is a visual interface to bridge the gap between users' perceptive concepts and direct manipulation query languages.