25

Kuafu: Annotate and Search for Images on the Web

by **Zhonghui Xu**

Bachelor of Engineering, Computer Science and Technology
Tsinghua University, China, 1991
Master of Science, Computer Science
Institute of Computing Technology
Academia Sinica, China, 1994

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements for
the degree of Master of Science
at the Massachusetts Institute of Technology

September, 1997

Author
Program in Media Arts and Sciences
July 25, 1997

Certified by
Ken Haase, Associate Professor of Media Arts and Sciences
Program in Media Arts and Sciences
Thesis Advisor

Accepted by
Stephen A. Benton, Chair
Departmental Committee on Graduate Students
Program in Media Arts and Sciences

Kuafu: Annotate and Search for Images on the Web

Zhonghui Xu

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements for
the degree of Master of Science
at the Massachusetts Institute of Technology

July 25, 1997

Abstract

Digital imagery has become commonplace because the World Wide Web has been growing rapidly. In order to facilitate global sharing of digital imagery, the goal of this work is to build a Web-based system, Kuafu, for image annotation and search.

Annotation and search in Kuafu is based on the content of images. In this work, a cognitive search model is proposed, and the visual and semantic content of images are distinguished. The cognitive model explains that people begin to search using semantic content, and find the right picture by judging whether its visual content is a proper visualization of the semantics given. Kuafu emphasizes semantic content in particular, but utilizes visual content as well.

Enabling users to query efficiently is a central role of an interface for image retrieval systems. The disadvantage of the query-by-example paradigm, which is used in most image retrieval systems today, is that the user has to have examples at hand before searching. In contrast, the query-by-design paradigm is proposed, where the user is able, first, to semantically express what s/he wants and, then, gradually specify visual content as designs. Kuafu supports the query-by-design paradigm.

Kuafu utilizes iconic descriptions similar to those in Media Streams because the iconic descriptions are good at representing semantics and are also able to represent the visual content of images. At present, Kuafu is experimentally running on a database of the American Press photos.

Thesis Advisor:

Kenneth Haase, Associate Professor of Media Arts and Sciences

This work was sponsored by the News in the Future Consortium

Kuafu: Annotate and Search for Images on the Web

Zhonghui Xu

The following people served as readers for this thesis:

Thesis Reader

Ronald L. MacNeil, Principal Research Associate
of Media Arts and Sciences
MIT Media Laboratory

Thesis Reader

Marc Davis, PhD
Member of the Research Staff
Interval Research Corporation

Acknowledgements

It is a pleasure to thank my thesis advisor, Ken Haase, for his guidance and support. Ken has given me the freedom to explore, and allowed me to share his thoughts with him. It was Ken who, in a research discussion, first conceived of the idea to use icons to search for images on the Web. Sincere thanks go to my thesis readers Ron MacNeil and Marc Davis who have both provided helpful critiques of, and suggestions about, this work.

I gratefully acknowledge Steve Swaldman's help with Java programming; discussion with, and inspiration from, Joel DeVries; and Mori Ken's experience with FramerD. Credits should go to Golan Levin for creating most of the icons used in Kuafu and Alan Ruttenberg for converting them into GIF format. My appreciation goes to Walter Bender and Daniel Gruhl for providing the American Press photos on which Kuafu is experimentally running.

Thanks to Janet Cahn and Michael Best for their close reading of the proposal and Warren Sack for discussion about the name of the system and for demoing Media Streams for me. I owe many thanks to Sawad Brooks, Karrie Karahalios, Jalia Yang, Laurie Hiyakumoto, Sara Elo, Pushpinder Singh, and Martin Hadis as well.

Walter Bender and the News in the Future Consortium have provided nutriment and support for this work, which are gratefully acknowledged.

Thanks also to Matthew Trunnell, Jon Ferguson, and Fred Martin for their help during the developing of Kuafu.

Finally, thanks to my parents, who themselves received little education but set me on this path, and to my sister and brother, who always support my every struggling step.

Table of Contents

List of Figures

Chapter 1

A scenario and motivation

Imagine that a little boy at home in France receives a phone call from his parents, who are traveling on business in the United States. The parents mention to their son that the foliage in the White Mountains of New Hampshire is tremendously beautiful. The boy, we can imagine, will be so impatient to see how beautiful it is that he cannot wait until his parents come back. He may want to search the Web, which is likely available to his family (or will be in the near future). However, search engines like Yahoo and Excite use English keywords that he may not know. On the other hand, in an icon-based search system, he may be able to create a semantically structured iconic query which does not rely on his knowing English, but rather on his understanding the iconography of the domain of images he wants to find. For example, he may be able to understand a geographical icon of the USA, an object icon of trees, and a time icon of fall because these icons are fairly intuitive and their meanings, in certain categories, are easy to guess. Therefore, the little boy may be able to find images of fall foliage in the USA, which other users have already annotated and associated with the search engine.

In the French boy's scenario above, the user has trouble making good use of today's search engines (Yahoo, Excite, etc.) that are based on English keywords. As the internet expands into every corner of the lives of average people in the world, language barriers are becoming problems that users of the internet have to face. This problem is serious even in Europe, where the countries are getting close to having unified markets and policies and official common languages such as English, German and French, but where people keep using their original languages in daily life.

The internet makes it possible for users all over the world to easily share information. Yet, global sharing of information ought to be less dependent on any single language. In the French boy's scenario, the user searches for an image rather than a piece of text. If there were a service from which digital imagery could be retrieved without the use of textual information, the language barrier would be avoided.

The goal of this work is to build a Web-based system for both image annotation and search with icons in order to facilitate global sharing of digital imagery. The system is called Kuafu, which is the name of the hero in a Chinese legend. The legend says that Kuafu chased the sun for ten days and died of thirst.

Chapter 2

A cognitive search model and query paradigms

Keyword-based matching is a basic operation in the process of textual information retrieval. Unfortunately, images are largely unstructured raw sensory data. Although the research about extracting visual features and high level structures from raw sensory data has been making progress [6, 23, 24], it is still far away from achieving the automatic annotation of images. Therefore, it is important to understand how people search for images, what kind of image content is significant at different steps of the search, and what kind of service would help the users search in a natural way. This section discusses these issues through the scenario of searching a family photo album, proposes a cognitive search model and a new query paradigm, and addresses related work in the literature.

2.1 A cognitive search model

Imagine that you have a family photo album at hand and want to find a picture taken at your fifth birthday party. What would you do? Can you give a detailed description of the picture (before you see it)? Can you remember what clothing you wore? What color? What texture? No. You most likely cannot. Maybe you would say "I (as a kid) should be in the picture and a birthday cake should be beside me." Yes, you are right. But you would still have difficulty specifying what *you* looked like at 5 years old and what the birthday cake looked like. That is, it is hard to give specific visual clues as a basis for a search.

In reality, you most likely just flip through the photo album and, suddenly, find the one that you want because you see *yourself* and *the cake* in the photo.

Let us divide this process into two steps. First, you begin to search pictures for the child, *yourself*, without specific visual information. However, you know that a child and a birthday cake should be in the picture, and you have "common sense" notions about the child and the event in your mind. For example, the child should be short, small, look excited, etc., and the birthday cake should have candles. Second, seeing a picture, you quickly judge whether it contains these "common sense" features. If it does, you recognize the child and the cake and have found the right picture. If not, you continue to another picture.

In the second step, it is important to notice that the judgement is not really made by "matching computation"[1]because the "common sense" notions are too elusive to be used directly for "matching computation." Rather, we think that the judgement is made based on whether you succeed in visualizing the "common sense" in your mind with what you see in the picture. Therefore, from the cognitive point of view, search is a process of "common sense" visualization.

How human brains visualize "common sense" remains a complex mystery. Yet, it is clear that the "common sense" visualization of one object can be different from one time to another. That's why you can recognize the same child from picture to picture despite changes in hair style, complexion, and clothing. We gradually obtain "common sense" about one thing (say the child) while seeing it many times under different circumstances. Thus, matching computation cannot capture the variability of the visualizations for "common sense" given.

[1]Matching computation of low level visual features is usually used as a technical basis of image recognition.

2.2 Semantic content and visual content

"Image content" is not a well-defined term, although it has been used in research. In the computer vision community, content-based retrieval is simply synonymous with feature-based retrieval [4,16]. In systems like IBM's QBIC [7], "content" is visual properties, such as shapes, colors, textures, and spatial relationships. In this work, image content is divided into semantic content and visual content. Their distinction is discussed below.

Based on the analysis of the search-photo-album scenario, we divide image content into semantic content and visual content. Semantic content is defined as "common sense" (knowledge) that we would have in the first step, and visual content as the specific visual features (properties) that would be found in the second step. The semantic content can be meanings, associations, and identifications, while visual content is strictly perceptual. The semantic content can sometimes have a high level structure like context (for example, waving to each other means greeting), but visual content cannot. Semantic content is abstract and fuzzy, referring to concepts of objects and situations, while visual content is concrete. Another distinction is that semantic content is qualitative whereas the visual content is quantitative. For instance, in the birthday party scenario, a cake with candles is thought of as a birthday cake no matter whether the cake is a cylinder or a cube, and no matter what size it is, where the birthday cake is semantic content but its shape and size are visual content.

According to the cognitive search model, people begin to search using semantic content, and find the right picture by judging whether its visual content is a proper visualization of the semantic content. From this point of view, two issues are raised. First, it is necessary that an image retrieval system concern itself with semantic content, as well as visual content which is generally much less "meaningful." A disadvantage of the systems that do not utilize semantic content is addressed in Section 2.4. Second, how can the user represent semantic content? How can the user query semantically? Section 2.3 discusses this issue in particular.

2.3 Query-by-design

Enabling users to query efficiently is a central role of an interface for content-based retrieval systems. It is important to provide services such that users can give "semantically meaningful" queries in a natural way. A query paradigm called "query-by-design" is proposed in this section. Comparison to other paradigms will be discussed in section 2.4.

Query-by-design means that the user can give queries in a way similar to design. Design is an approach to express semantics (concepts, feeling, meaning, impression...) with visual features. When one designs, s/he often draws a sketch first that looks like what s/he wants in the layout, and then modifies the sketch carefully. The sketch serves as semantic content because the sketch is not exactly the final result but represents the designer's rough idea.

The query-by-design paradigm has an advantage in that the user is not required to specify the visual content of images in the beginning of the search. Yet, the user can refine the search by adding to, or modifying, the visual content while keeping the semantic content.

From the technical point of view, it remains a hard job for computers to recognize semantic content from sketches [7]. In addition, it is inefficient for users to draw a sketch as a query. A practical solution is to provide a pictorial query language in which users can approximately express the semantic content of images they are looking for. This is one of the main reasons why iconic descriptions are utilized in Kuafu.

2.4 Related work and discussions

Clearly, managing and manipulating image data as images is not a trivial problem. What we really need to do is to build efficient content-based retrieval systems to help users search large databases of images in a natural way. Research projects about content-based retrieval systems have been going on in the past few years. The most comprehensive research projects include IBM's QBIC [7], MIT Media Lab's Photobook [4] and Media Streams [2, 36], ISS' SWIM [21], and UNM's CANDID [9]. QBIC, Photobook, and Media Streams are discussed below.

QBIC (Query By Image Content) allows queries of large image databases based on visual image content – properties such as color percentages, color layout, and textures occurring in images. QBIC also supports retrievals based on a rough user sketch, which is used just as a means to input visual properties but not semantic content as in query-by-design [7]. The Photobook system is a general tool that can be tailored for a number of applications to content-based retrieval. Photobook works by comparing features (parameter values of particular models such as color, texture, and shape) associated with images, not the images themselves [4]. Thus, image content utilized in QBIC and Photobook really refers to visual content in the context of this thesis. One disadvantage is that they have little ability to query the semantic content of images. So a query using a red car on a green background will possibly match a red bird on a similar green background [7].

In addition, QBIC and Photobook essentially adopt the query-by-example paradigm. That is, the user provides image examples (or properties, features, etc.) as queries, and images that are similar to those examples are retrieved. The query-by-example paradigm has three disadvantages. First, visual features in image examples are too concrete to represent general semantic content. Second, image examples are themselves just images and hard to compose into complex queries. For instance, given two people's pictures (as examples) separately, the user has difficulty expressing the people shaking hands.

Third, in reality, when a user wants to search for something, s/he usually does not have an example at hand. In cases like the search-photo-album scenario, the user is not even able to give specific visual properties or features.

Media Streams uses a stream-based, semantic, memory-based representation with an iconic visual language interface to annotate video for content-based retrieval. Media Streams uses semantic content and visual content and supports query-by-description as well as query-by-example. Query-by-description is a general approach using some query language, which may be SQL; a natural language, or a restricted subset of one; or, as in Media Streams, an iconic visual language. Whereas, query-by-design, from the cognitive point of view, tells how the user ought to query about images. That is, query semantic content first and then visual content.

In conclusion, we think that an efficient image retrieval system should use both semantic content and visual content. Because people are generally better at dealing with semantic content rather than with visual content, the query-by-design paradigm would be a good approach by which the user can query in a natural way.

Chapter 3

Icons and an iconic language

The word "icon" derives from the Greek *ikon*, which implies a mode of communication using primitive visual imagery that relies on the ability of people to perceive natural form, such as shapes and motion [31]. An interesting example of an iconic message is the interstellar greeting card attached to the Pioneer 10 space probe, where there were no words, but rather only the graphics of a man, a woman, and other visual symbols (figure 1).

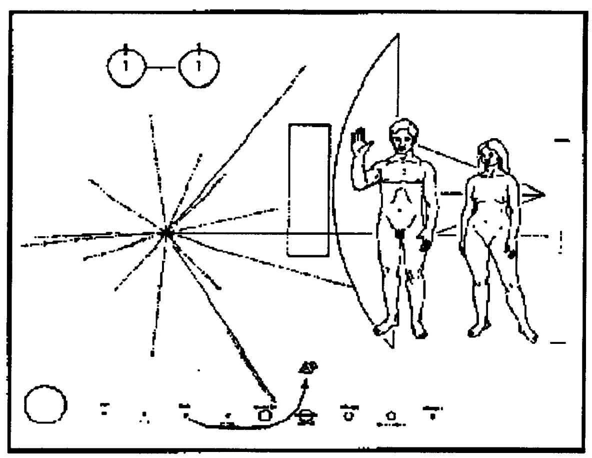


Figure 1: The Pioneer 10 spacecraft's interstellar greeting card.

For centuries, researchers have made efforts to create iconic languages to facilitate global communication, e.g., C. Bliss [27] and O. Neurath [32]. More recent are S. Lee's 2D C-string, which allows for the natural construction of iconic indexes for pictures [17, 43], M. Davis's Media Streams system, which uses icons to annotate video content [2, 36], and S.K. Chang's BookMan, which is an iconic interface to a virtual library [15]. Of them, Media Streams is novel in that its icons have a semantic status different from icons used in traditional graphical user interfaces: the objects and processes denoted by Media Streams' icons are not computational ones, but aspects of the video content which they are designed to represent.

Similar to Media Streams, Kuafu uses icons to represent the content (semantic and visual) of single, still images, and adopts the icons[2]and many of icon structures used in Media Streams. This chapter addresses the advantages, meanings, and organization of icons, and points out a practical difference between Kuafu and Media Streams in section 3.5.

3.1 Why icons?

The advantage of icons lies in that they are visual and imply more information than keywords of the same meaning. For instance, seeing a house icon, the user can understand what it refers to and, in the meanwhile, can become aware of common visual features about houses. Visual features are usually hard to describe in a natural language but, as visual clues, help the user greatly to think graphically and understand what s/he is thinking about. That is why iconic signs (i.e., road signs and business signs) are commonly used in our daily life.

Kuafu uses an iconic representation because iconic representation is suitable for representing the semantic content and visual content of images, both of which ought to be explored in an efficient image retrieval system, as discussed in chapter 2. Using iconic representation is also an ideal approach for avoiding the language barrier that global internet users are faced with.

Further, as addressed in section 3.2, the icons in Kuafu are organized in semantic hierarchies so that inheritance relationships between icons

[2]Most of the icons used in Media Streams [2] and Kuafu were created by Golan Levin. Alan Ruttenberg made great contribution to converting the icons in ResourceEdit format on Macintoshes to GIF format. I created a small number of icons for Kuafu.

are implied. This is especially useful for search. Section 4.4.2 gives an example of this, and the server is capable of utilizing such inheritance relationships to retrieve images responding to users' queries.

3.2 Organization of icons

Kuafu's icon catalog includes human beings, objects, spatial relationships, sports objects, text characters, time, locations, weather, and colors. The catalog occupies a special space in the icon database and may change from application to application (see section 4.3). Figure 2 shows the current catalog, which includes 11 categories.

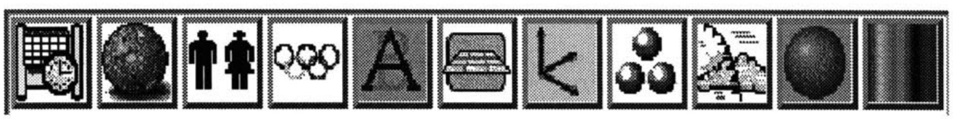


Figure 2: The current icon catalog in Kuafu includes 11 categories.

An icon category is called an icon family, which is organized in a hierarchical structure and displayed as an icon cascade. In an icon hierarchy, any two icons from different rows are semantically compatible but any two from one row are not. Some of the icons are shadowed, meaning that there are subfamilies of icons under them. Click on the shadowed icons, and the icons of their subfamilies will show up.

For example, figure 3 shows the family of time icons. Vertically, the family has year icons, season icons, and day/night icons and, horizontally, each row has icons of specific years, seasons, or times. Any two icons from different rows, say the day icon and the spring icon, are semantically compatible but any two from one row, say the spring icon and the winter icon, are not compatible. The day icon and the night icon have shadows because there are subfamilies of icons of more specific times under them.

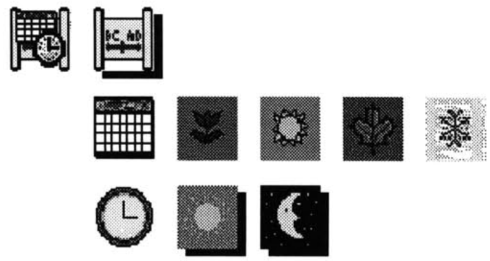
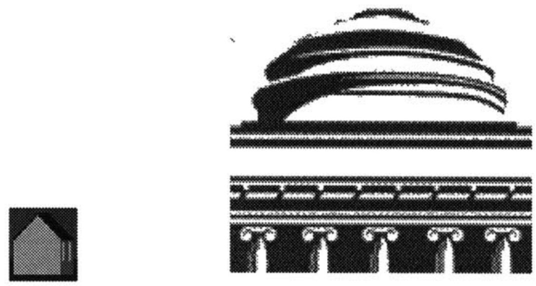


Figure 3: The family of time icons (not all the icons are displayed). The family has year icons, season icons, and day/night icons in three rows. Any two icons from different rows, say the day icon and the spring icon, are semantically compatible but any two from one row, say the spring icon and the winter icon, are not. The day icon and the night icon are shadowed. Click on them, and the icons of more specific times will show up.

3.3 Meaning of single icons

Individual icons in icon hierarchies are called single icons. The meaning of a single icon is twofold. First, it represents whatever looks like it (same or similar). That is, we use it directly according to its visual content. Second, an icon abstractly represents a class of objects, relationships, properties, and so on and so forth, with the objects' semantic contents as bases. That is, we need to recognize what an object is and then use it to represent whatever is associated to it. At this level, we don't care much about how different an icon looks from a real object.

For instance, a building icon in figure 4(a) can be chosen to represent a building with a dome in figure 4(b), although they look quite different. Note that the building icon represents houses and buildings that may look different, and have different visual contents. We can certainly create a category of building icons to represent a variety of buildings in detail so that the difference in appearance between the icons and the real buildings will be decreased. Media Streams actually does in this way. But keep in mind that it is impossible and unnecessary to create an icon for every building in the world.



(a) (b)

Figure 4: Meaning of a single icon. An icon semantically represents a class of objects that may look different from one another. Thus, the icon in (a) can be used to semantically represent the building in (b).

3.4 Compound icons

The single icons in Kuafu can be combined to generate compound icons according to the users' needs. Compound icons are important because they can represent diverse and rich semantic content that users desire. Another advantage is that the number of compound icons that can possibly be made from a small set of single icons is far larger than the total number of the single icons.

Kuafu has two types of compound icons[3], as in figure 5(a) and (b), respectively. The compound icons of the second type, in figure 5(b), have a syntax similar to that of natural languages.

Kuafu contains nested compound icons[4]as well. The nested compound icons have more complicated meanings than do the compound icons. Figure 6 shows an example of a nested compound icon with the meaning that a basketball in a basket is in the upper-middle position

[3] In Media Streams, compound icons and glommed icons are strictly distinguished, corresponding to figure 5(a) and (b), where the elements of a compound icon are all from the same hierarchy and the elements of a glommed icon are from different hierarchies[2].

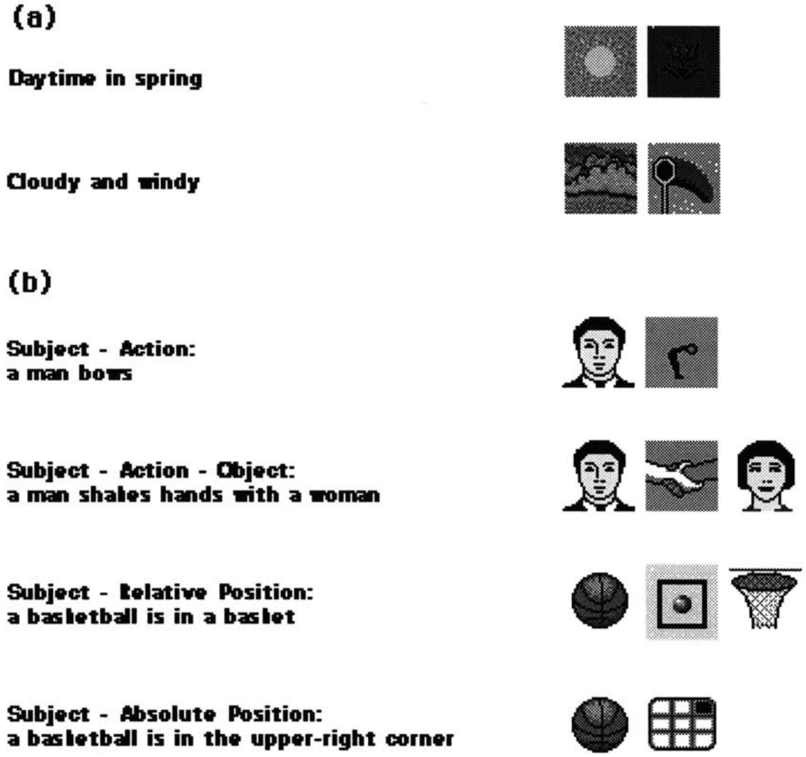


Figure 5: Examples of compound icons

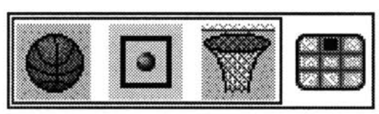


Figure 6: A nested compound icon, meaning that a basketball in a basket is in the upper-middle position of an image.

of an image.

How to generate compound icons and nested compound icons will be discribed in chapter 4.

[4] Media Streams does not have nested compound icons.

3.5 Changeable icon catalog

A difference worth noting is that Kuafu's icon catalog is changeable and can include categories of objects for domains of special interest. Media Streams also has many subcategories under the object category. However, a special interest subcategory in Media Streams could be brought up to the top level catalog in Kuafu. The reason this is an advantage is practical rather than theoretical. There are so many kinds of objects in the world that their hierarchy is quite deep, and finding icons in a deep subcategory will be time-consuming. Therefore, placing a special interest subcategory in the top level catalog will bring convenience.

Experimentally, Kuafu is running on a database of the American Press photos, most of which are sports news photos. Thus, Kuafu's current catalog particularly includes the icon category of sports objects. See figure 2.

Chapter 4

Implementation of Kuafu

This chapter describes the components and implementation of the Kuafu system[5]. Section 4.1 gives an overview of its architecture and section 4.2 depicts its user interface. The database management is described in section 4.3, and the algorithms for the annotation and search of the server are provided in section 4.4.

4.1 The architecture

The Kuafu system consists of a Web-based interface, an annotation and search server, and a database management system, as shown in figure 7. Through the interface, a user can annotate his/her own images online and/or search for interesting images that other users have already annotated. There are four supporting databases containing icons (K-icons and V-icons, see section 4.3), URLs, and indexed images, respectively. Note that Kuafu does not store annotated images but rather their indices and URLs.

While the interface is written in Java, the server and the database management system are written in a combination of FramerD and Java. FramerD is a persistent framework supporting cross-platform knowledge representation and database functionality [1]. The reason to use FramerD lies in its efficiency and scalability.

[5] The old name of the system is Iconish.

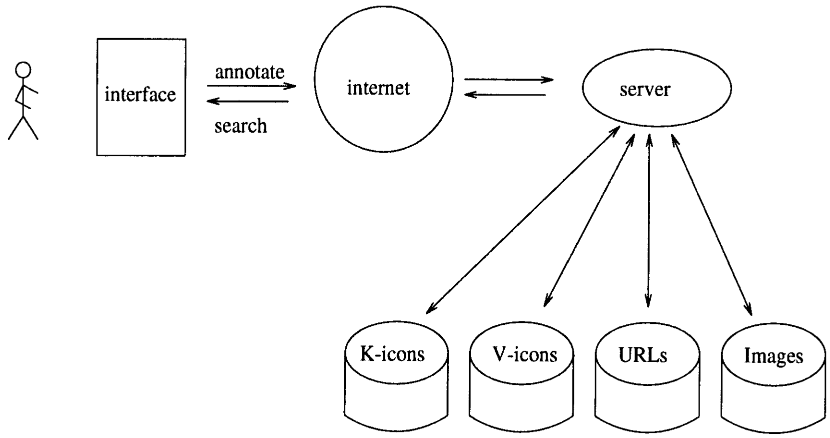


Figure 7: The architecture of Kuafu: Kuafu consists of a Web-based interface, an annotation and search server, and a database management system.

4.2 The interface

The interface has two working modes, the annotation mode and the search mode, as appears in figure 8 and figure 9, respectively. The user is able to annotate images in the annotation mode and search in the search mode. In either mode, the interface is composed of the icon space, icon pool, icon palette, icon board, prompt area, and display area.

In the icon space, an icon catalog and open icon hierarchies are shown. The user can browse all single icons organized in a hierarchical structure, and select any of them. In the icon pool, compound icons the user has created are saved so that they can be used for queries, annotations, and composing nested compound icons. The compound icons in the pool are also organized in a hierarchical structure according to their first icons and displayed vertically. The icon palette is a place where compound icons are composed, and the icon board temporarily displays icons (single and/or compound icons) for queries or annotations. Figure 10 shows icon flows to the icon board from the

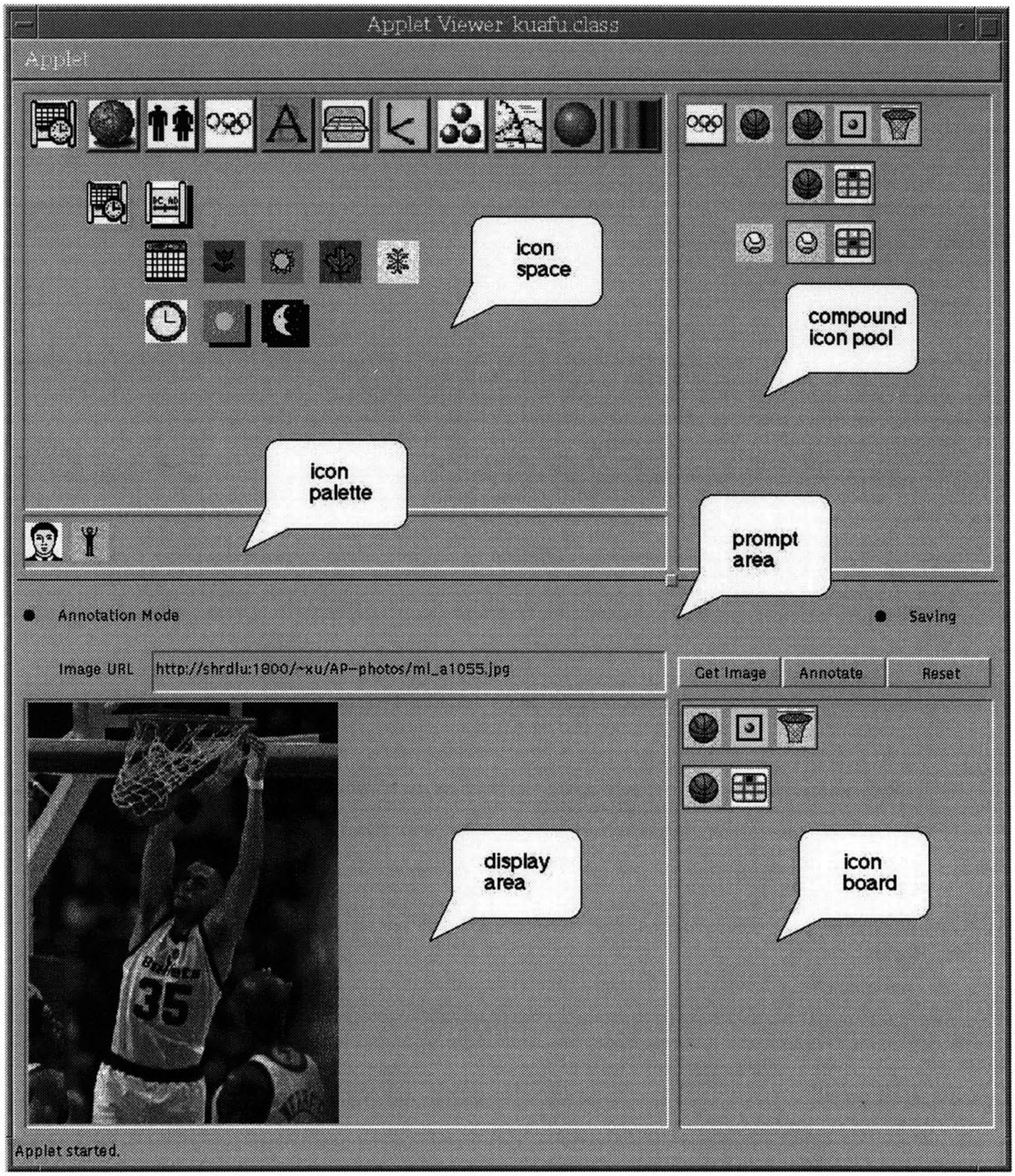


Figure 8: Kuafu's interface in the annotation mode. The interface is composed of an icon space where a main icon catalog and an open icon hierarchy are shown, an icon palette where single icons are combined to generate compound icons, a compound icon pool where compound icons are saved, an icon board where icons are given as annotation, a prompt area where commands and prompts are shown, and a display area where an image being annotated is shown. Note: the compound icons saved in the icon pool are also organized in a hierarchical structure.

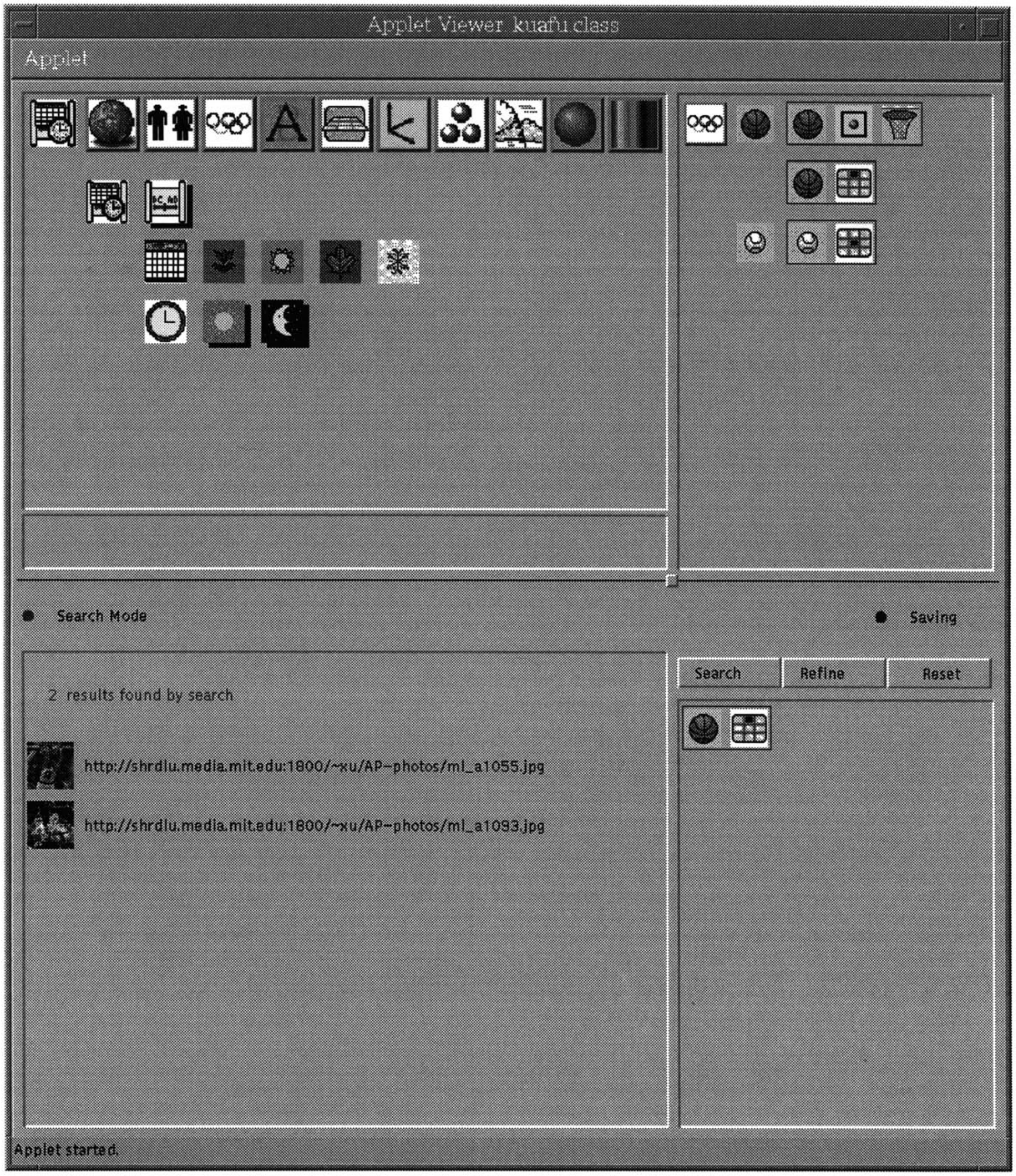


Figure 9: Kuafu's interface in the search mode. The interface basically looks similar to that in the annotation mode. The differences are 1) the icons in the icon board serve as queries; 2) search results are shown in the display area; 3) there is no image URL text field in the prompt area.

icon space, from the icon pool, and from the icon palette.

The display area shows images to annotate in the annotation mode or search results in the search mode. The prompt area shows command buttons and prompts to inform the user of what s/he is doing. There is a small, round mode switch (red/blue) at the right of the prompt area and a saving switch at the left. The mode switch changes Kuafu's working mode from annotation to search, or vice versa, and the saving switch indicates whether compound icons are being saved in the icon pool.

Chapter 5 will describe in detail how to use the functions in the interface for iconic annotation and search.

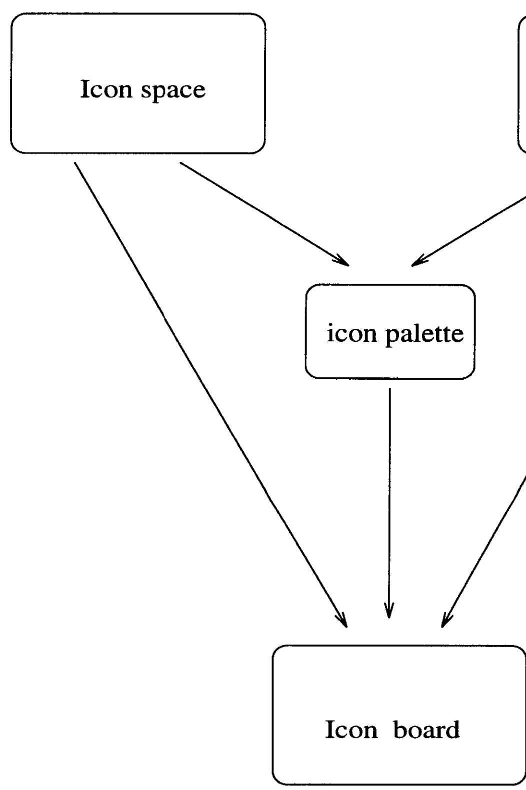


Figure 10: The icon flows to the icon board from the icon space, from the icon pool, and from the icon palette. See figure 11 and figure 12 as well.

4.3 The databases

Kuafu has a K-icon database, a V-icon database, a URL database, and an indexed image database.

The K-icon database stores all single icons that can be browsed in the icon space of the interface. The icons in this database are named K-icons (Knowledge-icons) because they are organized in a hierarchical, semantic structure that implies knowledge of them. In the K-icon database, the addresses from 0 to 127 are reserved for the icon catalog, one address per category. Kuafu's catalog can include 128 categories at most and, to add a category, it just needs to be placed between address 0 and 127.

The K-icon database is built in advance and will be downloaded to the interface (client) side when the user accesses the Kuafu site.

The V-icon (virtual-icon) database stores all icons used as image annotations. In contrast to the K-icon database, the V-icon database is dynamically formed while users annotate their images. Some of the icons in this database are single icons and others are compound icons. "Virtual" refers to the fact that the compound icons are created by users so they cannot be included in the K-icon database. When the user creates a compound icon and uses it to annotate an image, this compound icon will be stored in the V-icon database, if it is new. Thus, the V-icon database may grow as the number of annotated images increases.

The URL database serves as a hash table in order to give a unique URL object to every URL address. The URL objects are used as the internal representation of the URL address, and as references in the indexed image database. The advantage of using the URL database is a saving of disk space and an increased efficiency.

The indexed image database is the major working database with which the engine searches for desired images. Importantly, it is not necessary to store images locally, but only their indices, since the images are indexed by icons.

The V-icon database, URL database, and indexed image database all stay on the server side. They are not downloaded to the interface side in order to save time.

It is clear that the database management system has to be able to deal with very large databases because the World Wide Web is growing rapidly (although annotated images will not be stored locally). FramerD is an ideal database management system in that it provides a basic indexing mechanism to achieve high efficiency and scalability [1].

4.4 The server

As the core of the Kuafu system, the server communicates with the interface and implements all the functions in response to the user's queries. Its basic capability is to find appropriate icons and indexed images with help from the database management system.

One critical issue is that the server works on indexed objects. The databases do not really store icons, URLs, and images, but indexed icon objects, indexed URL objects, and indexed image objects. Here, "object" is a term in the context of FramerD [1].

Annotation and search are two major functions that the server provides. The following two subsections describe the algorithms of how the server operates on indices of the V-icon database, URL database, and indexed image database.

4.4.1 The annotation algorithm

To annotate an image, the user needs to fetch the image by giving its URL, and then generate an annotation by selecting appropriate single icons from the icon space, selecting compound icons from the icon pool, and/or creating new compound icons in the icon palette.

From the internal point of view, the server needs to: 1) index the URL in the URL database and generate a unique URL object; 2) index the iconic annotation in the V-icon database and get icon objects (and meanwhile add the icons in the V-icon database if they are new); 3) index the image with the URL object and the icon objects in the indexed image database.

While being indexed in the V-icon database, the iconic annotation is decomposed and components of compound icons are extracted. The image is indexed by icon objects of the components as well.

4.4.2 The search algorithm

While searching, the server first indexes an iconic query given by the user to obtain its indexed icon objects, then retrieves indexed images in the indexed image database using the icon objects as a basis, and finally refers to corresponding URL objects in the URL database.

It is important that another algorithm is evoked in the above process. This algorithm infers a relationship between the icons given in the query and the others in the K-icon database. Additional icon objects will be used for search if inheritance relationships between them and the icons in the query are found. This algorithm makes it possible that more semantically related images can be retrieved. For instance, if a user searches for images related to "USA" in a keyword-based system, he will only get URLs where textual annotations of images contain the word "USA." Yet, in Kuafu, search with a geographical icon of the USA will bring up URLs of images annotated by a geographical icon of Massachusetts. This result is helpful because Massachusetts is one state of the USA. In fact, this kind of knowledge is implied in the icon hierarchical structure in the K-icon database.

Chapter 5

Iconic annotation and queries in Kuafu

Kuafu is a tool to annotate and search for images on the Web, using semantic content as a basis. Both annotation and queries are iconic descriptions (without any text strings although there are text character icons provided). Kuafu adopts the query-by-design paradigm. Users thus can combine icons to express relatively complex semantic content in a way similar to design.

The appearance of Kuafu's interface is displayed in figure 8 and figure 9, in the annotation mode and search mode, respectively. This chapter describes how to generate iconic annotation of images and how to give iconic queries for search through the interface.

5.1 Working modes

Kuafu works in the annotation mode for annotation and in the search mode for search, as shown in figure 8 and figure 9, respectively. In order for users to tell easily what they are doing, there is a mode prompt and distinctive command buttons for each mode. The mode switch is the small, round button (blue/red) at the right of the prompt area. Before annotating or searching, users need to make sure that Kuafu is in the working mode they want, and to click on the switch to change if necessary.

5.2 Selection of single icons

To select a single icon, the user needs to open a proper icon hierarchy:

to look through the icon hierarchy, go to a deeper level of the hierarchy if necessary, find a desired icon, and click on it. The icon chosen will go to the icon board or the icon palette if the user is making a compound icon. See figure 11.

5.3 Generation of compound icons

As has been addressed in section 3.4, compound icons are composed of single icons or other compound icons that are already generated. The syntax of the compound icons is discussed in section 3.4. To generate a compound icon, the user needs to 1) activate the icon palette by clicking anywhere in it; 2) pick icons from the icon space and/or from the icon pool; 3) click the palette to make it inactive and end the process. As a result, the compound icon just made will go to the icon board, and to the icon pool if it is new and the saving switch is on (blue). The compound icons saved in the icon pool are also organized in a hierarchical structure. Once saved, the compound icons will automatically be added to the corresponding hierarchy in the icon pool.

5.4 Generation of queries

To generate a query, the user needs to pick single icons from the icon hierarchies, generate new compound icons in the icon palette and/or select saved compound icons from the icon pool, and drag them to the icon board. The results, which are a list of URLs of images found by search, will be displayed in the display area. Click on the URLs, and the entire images will be downloaded and seen.

While the search command leads to a search of the whole indexed image database, the refine command leads to a search based on previous search results. Thus, the user can add new icons to the query to refine the search results.

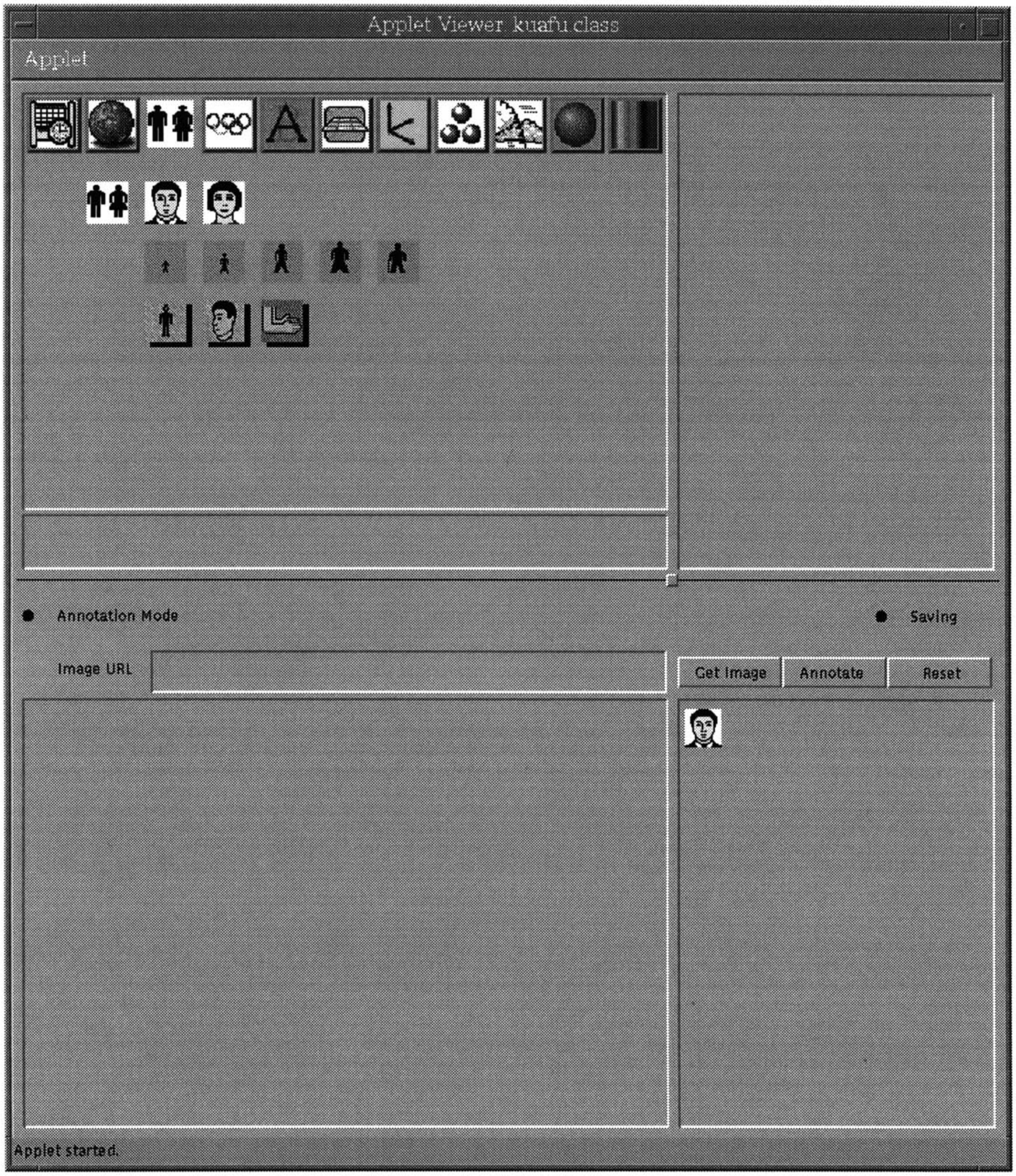


Figure 11: Pick a single icon in an open icon hierarchy by clicking on it. The icon will show up in the icon board, or in the icon palette if the user is making a compound icon. In this figure, the man icon in the human icon hierarchy has been clicked on and thus shows in the icon board as a result. See figure 10.

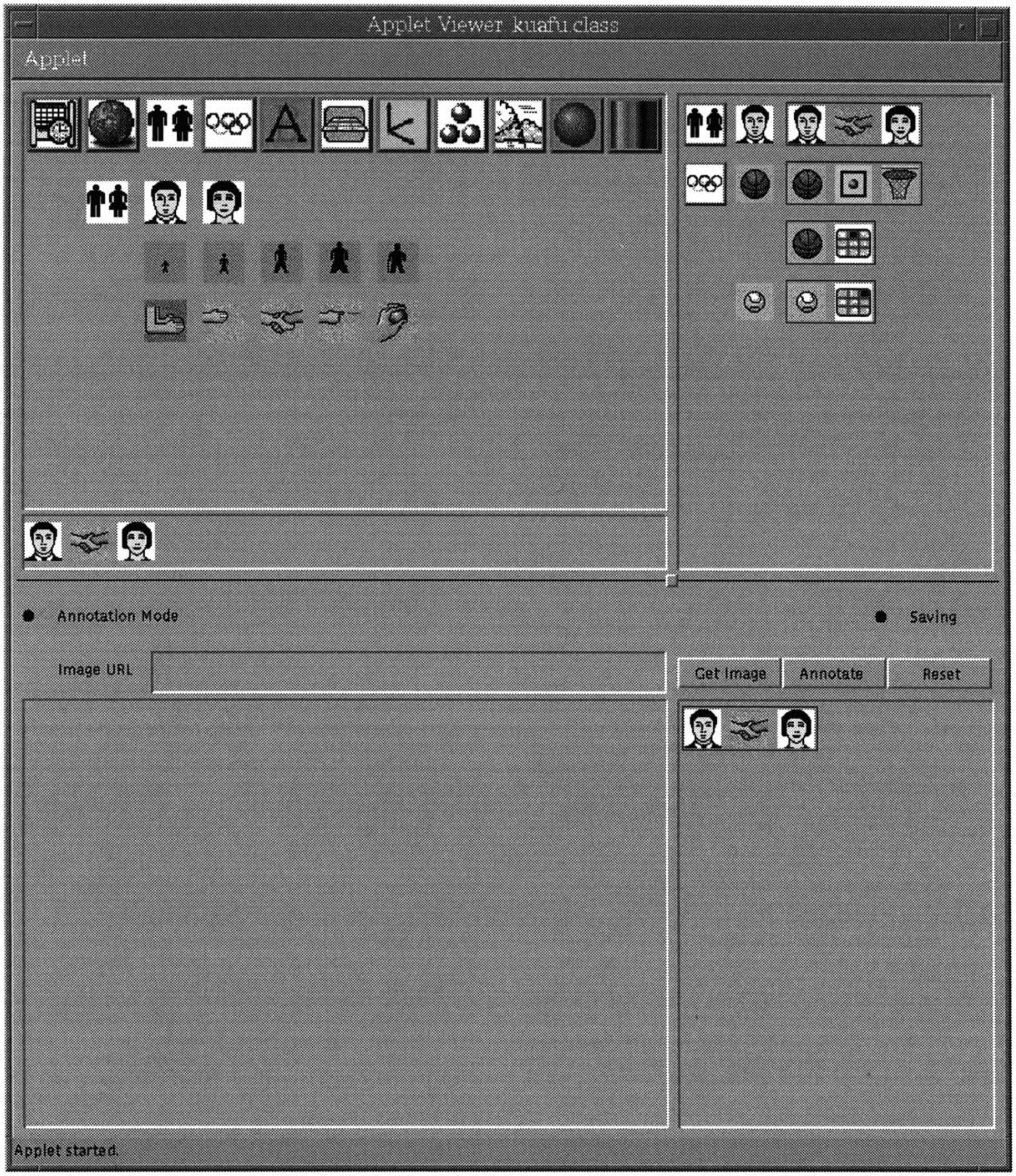


Figure 12: Generation of compound icons takes place in the icon palette. First, click anywhere in the palette to activate it; second, select icons (a man icon, a woman icon, and a shaking-hands icon) from the open hierarchy for the palette; third, click the palette again to end the process. Once a compound icon has been generated, it appears in the icon board, and in the icon pool if it is new.

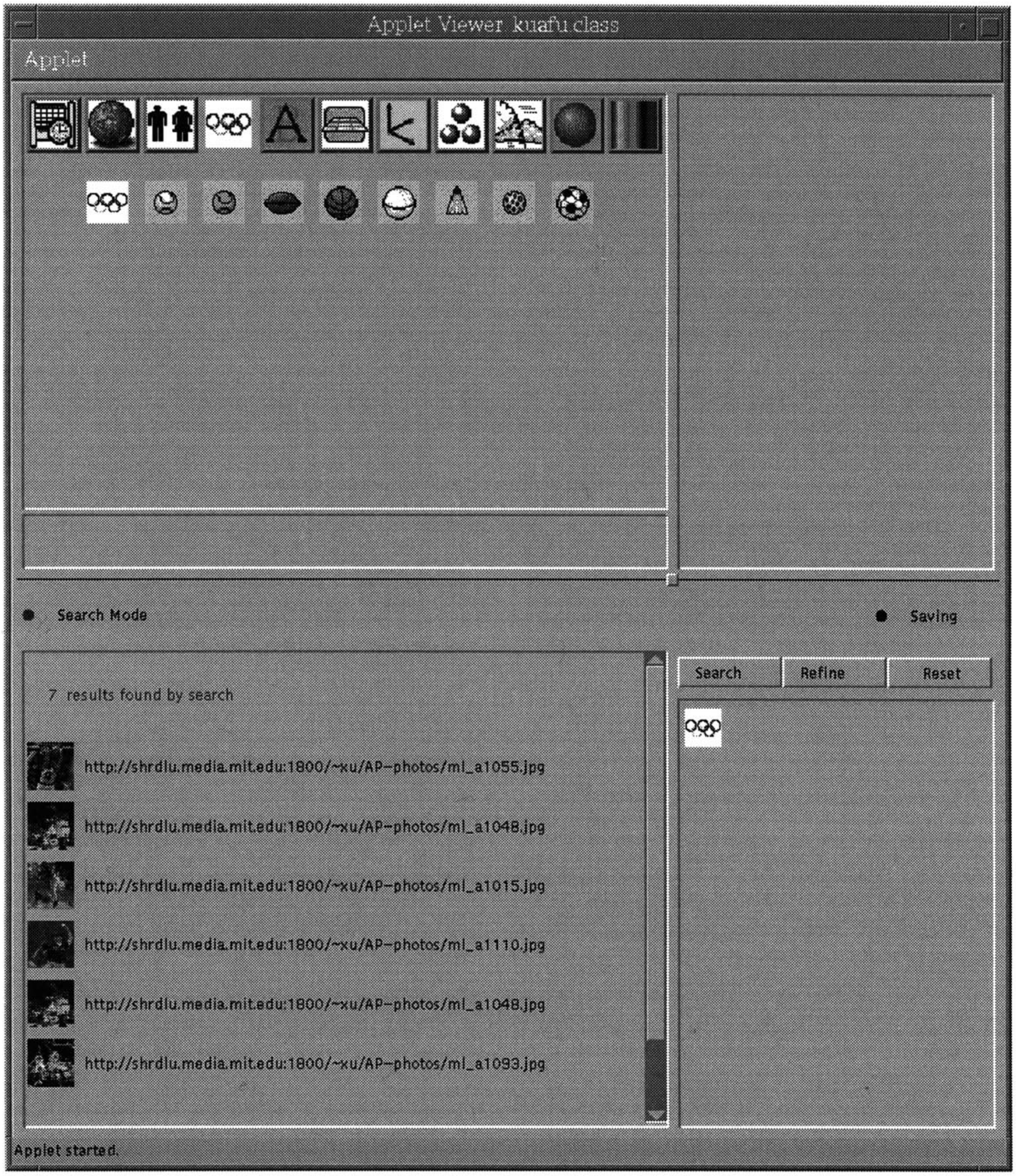


Figure 13: Search with icons given in the icon board. The icon picked as a query is a general sports object icon so all pictures with sports objects are retrieved. The search results are displayed in the display area. Notice that in the query-by-design paradigm, users first give generic queries and then detailed queries. In this example, the user can then query for a specific sports object, say a basketball, and afterwards query for relative details about the object, say the position of the basketball.

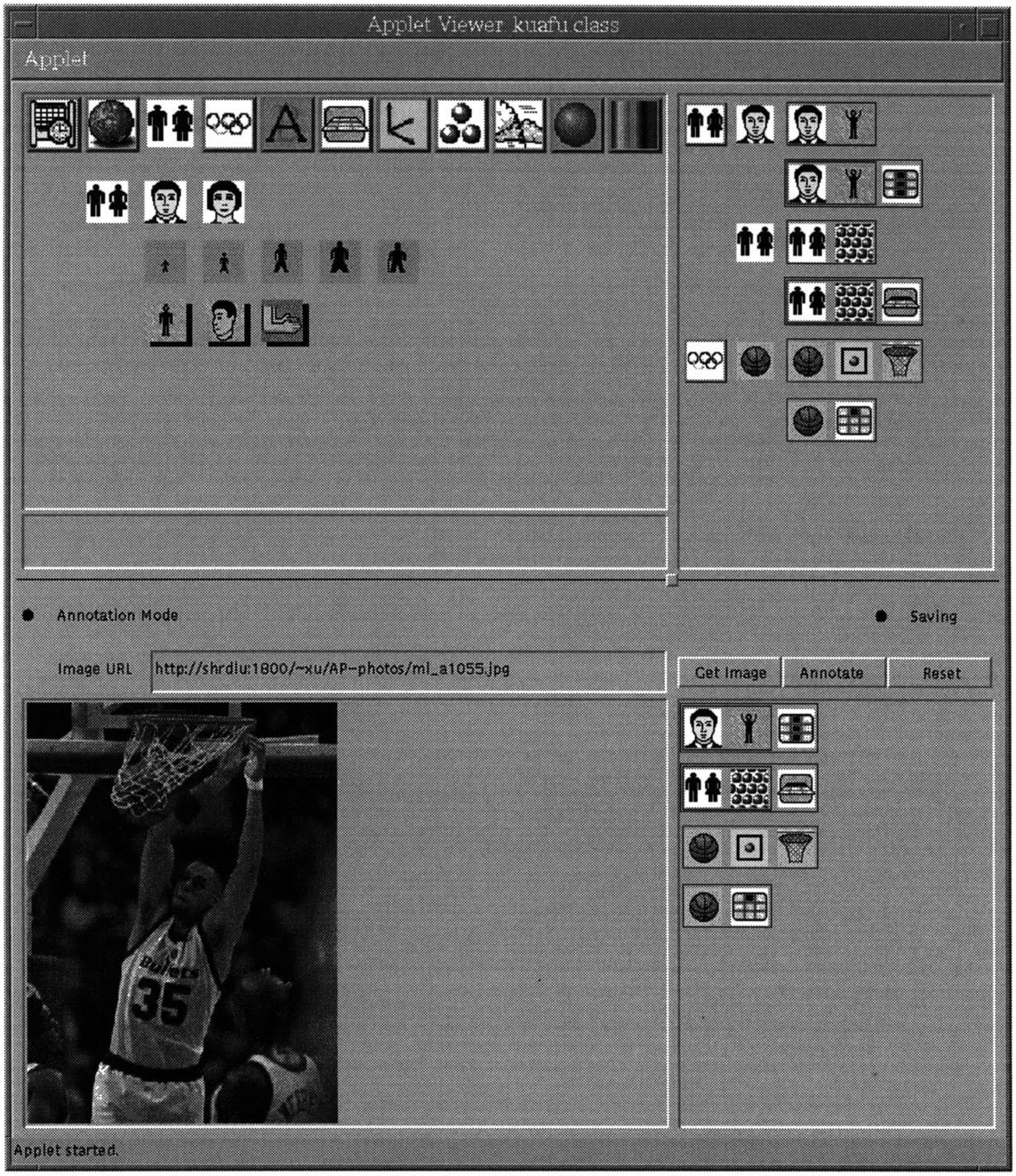


Figure 14: Annotation of the image in the display area. The picture shown in the display area is a sports news picture from the American Press photo database. Its URL is given in the image URL text field. Four compound icons are given in the icon board as its annotation. Note: the compound icons saved in the icon pool are also organized in a hierarchical structure according to their first icons.

5.5 Generation of Annotation

To annotate an image, the user needs to type in its URL in the image URL text field. The corresponding image will show up in the display area. Upon seeing the image, the user needs to pick single icons in the icon space, generate compound icons in the icon palette, and/or select compound icons in the icon pool for the icon board to represent the content of the image. Click on the ANNOTATE button to submit the annotation. Then, the information of whether the submission has succeeded will be displayed.

Chapter 6

Experimental running: American Press Photos

We have a continuous feed of American Press (AP) photos coming into the Media Lab for the News in the Future (NiF) consortium (about 1000 per day). Kuafu is not yet open to the public on the Web but is experimentally running on a database of the American Press photos. It was also demostrated at the NiF sponsors meeting in May, 1997.

What we have learned from the experimental running is that the photos are so diverse that we need to have more icons to annotate them. From this, we can know what kind of highly specific icons are necessary for annotating news photos. Currently, most photos in the database are sports news photos.

The following is one example of annotation of the photos in the database and one example of search.

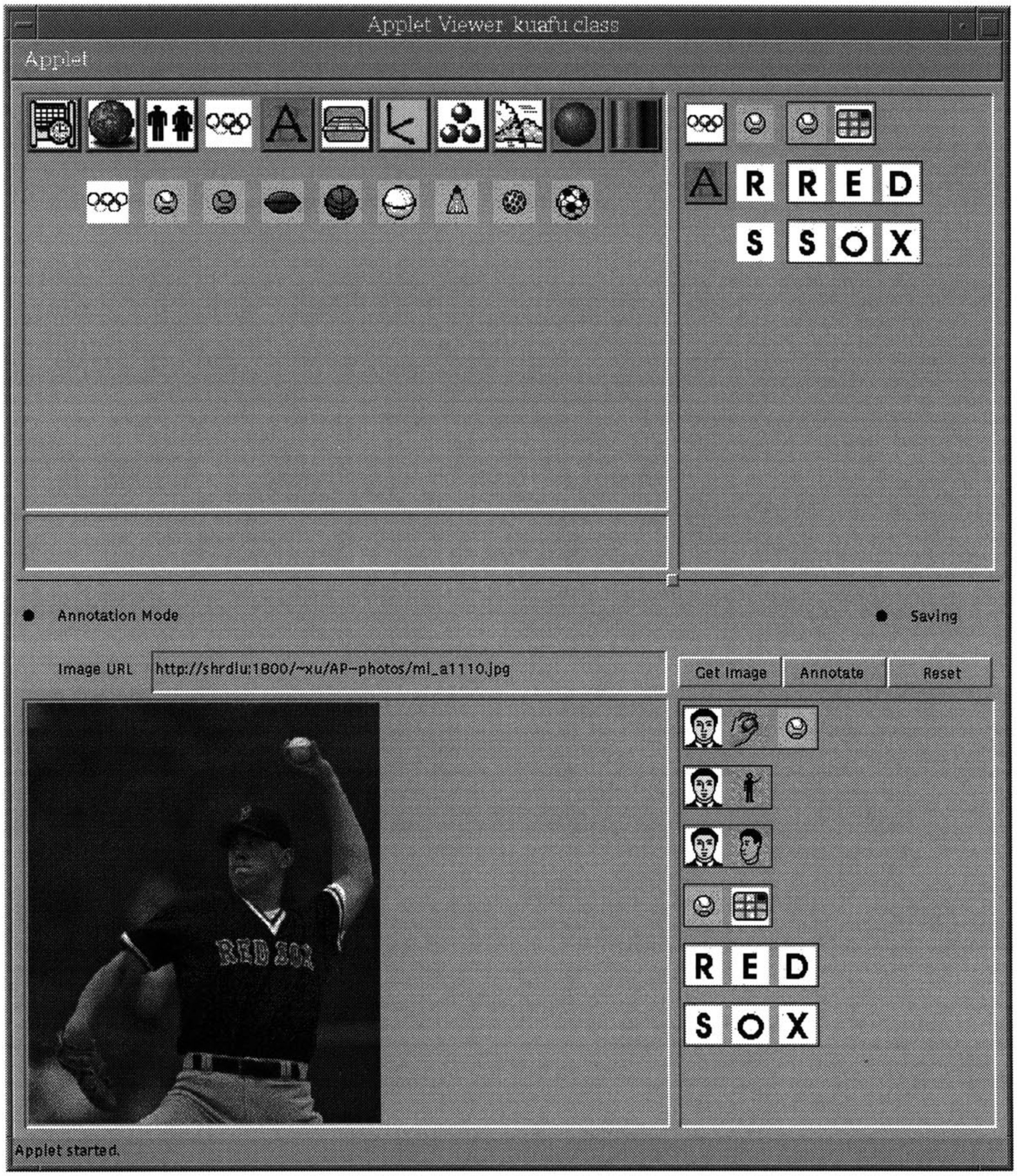


Figure 15: Example 1: annotate a picture of a baseball game. The picture from the American Press photo database is shown in the display area and its URL in the image URL text field. Its annotation includes the actions of the player, the position of the ball, and the name of the team.

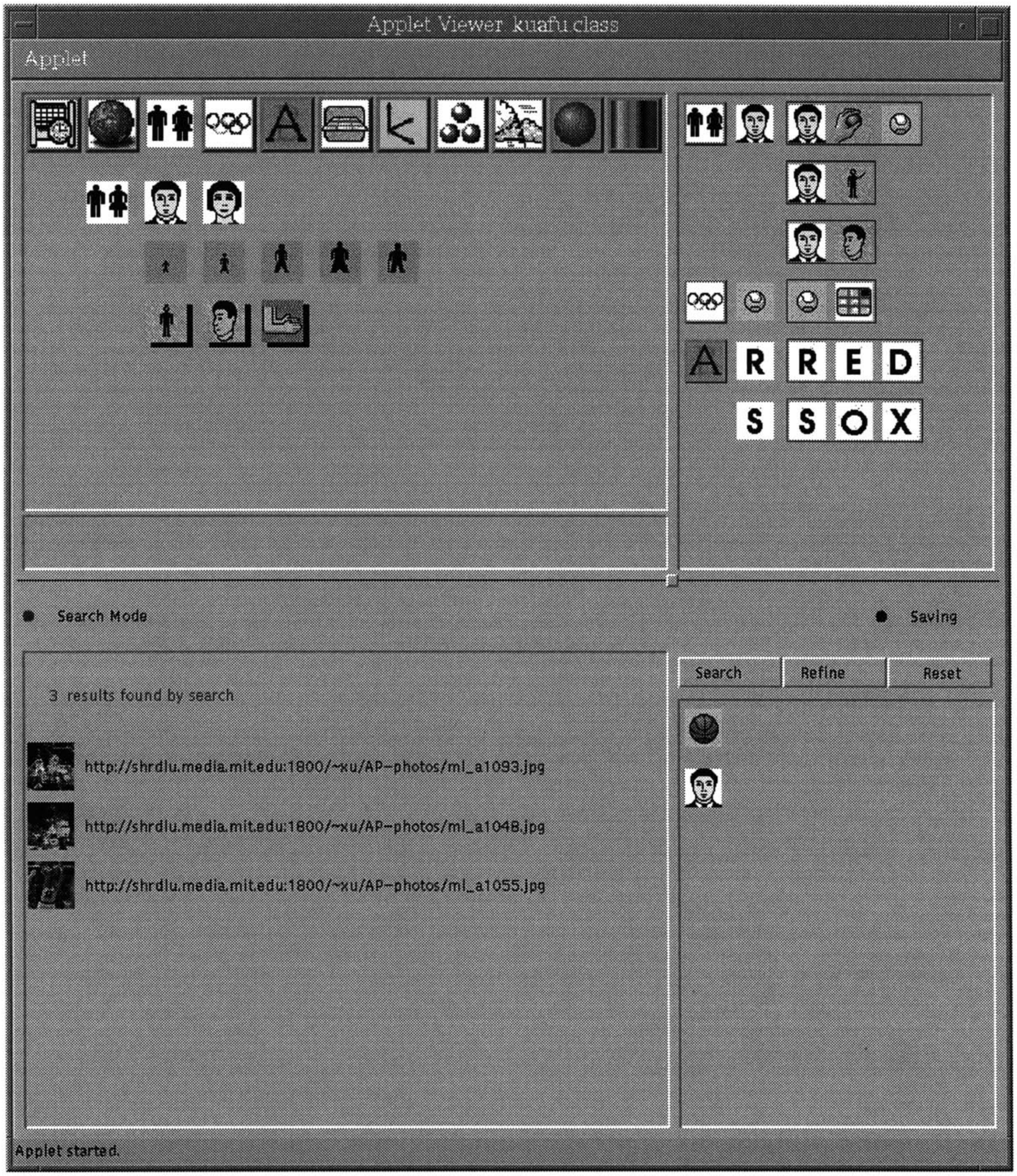


Figure 16: Example 2(a): search for pictures of men's basketball games. Three results are found and displayed in the display area. Click on the second highlighted URL and see the entire picture, as in figure 17.

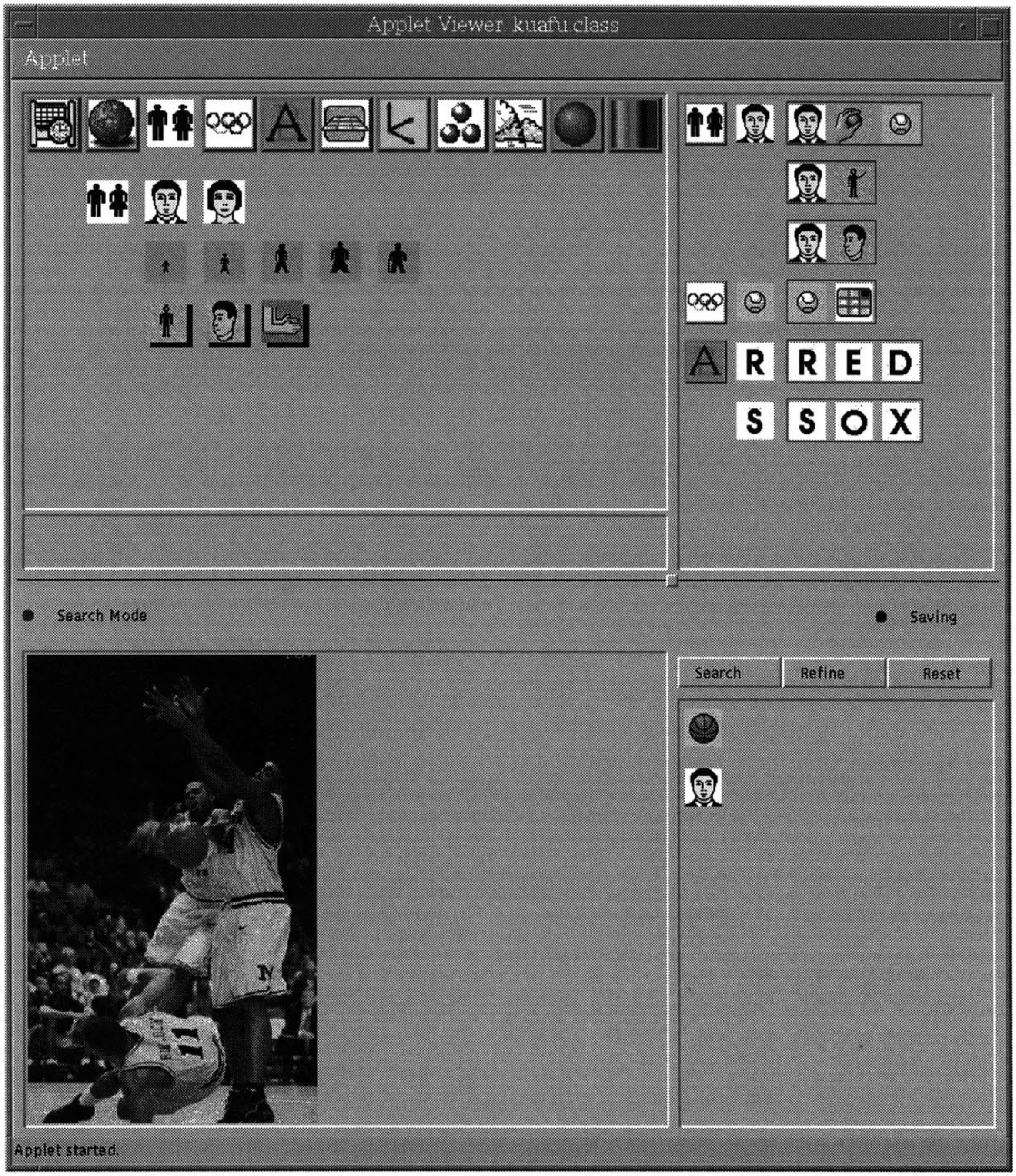


Figure 17: Example 2(b): search for pictures of men's basketball games. This is the entire picture corresponding to the second URL in the result list in figure 16. To go back to see the search results, click on the picture.

Chapter 7

Evaluation

7.1 Advantages and disadvantages

Iconic descriptions have advantages over textual ones in that they are visual and able to represent the semantic and visual content of images at the same time. It is a natural way to describe the visual image content of images with visual iconic primitives. It is especially important that such an iconic language can support the query-by-design paradigm very well.

Another advantage is Kuafu's scalability. Image retrieval based on visual feature matching is usually so time-consuming that scalability is a bottleneck with the retrieval systems of that kind whereas, using icons, Kuafu theoretically avoids the heavy load of the matching computation of visual features. In practice, FramerD, which Kuafu utilizes to manage the databases, provides an efficient indexing mechanism and is good at scalability. Importantly, this advantage makes it possible for Kuafu to deal with a growing number of images on the Web.

Nevertheless, there are two disadvantages. First, Kuafu doesn't have the ability to automatically extract "meaningful" image content from image data. Manually annotating images in a vast image database will be a tedious job. Second, Kuafu has no ability to annotate and search for low level visual features so, if that is what you are interested in, the results of search will be semantically, but not visually, specific.

Chapter 8

Other applications and future work

8.1 Other applications

The techniques developed for Kuafu can be applied to other fields. One possibility is to associate icons with keywords from multiple languages to help users like the French boy in the scenario in chapter 1.

In Kuafu, the semantic hierarchies enable general annotations for specific things. However, users may have difficulty finding specific things with general queries. As Professor Ron MacNeil pointed out reading the proposal, it would be good to archive images of a specific domain or to run the system on an intranet. In such a case, the icon catalog can be restricted to a specific domain and the icons can be more detailed so that the system will perform better.

8.2 Future work

Because automatic or semi-automatic annotation could save time in annotating a large number of images, and search based on low level visual features could refine search results, we hope that we can incorporate the techniques of automatic annotation, and search based on the low level visual features in images, in Kuafu in the near future. In addition, Media Streams allows for search based on the annotation of images that are found by previous search. Kuafu will implement a similar capability in a later version.

References

[1] K. Haase, "FramerD: A Persistent Portable Representation Library," Proceedings of the European Conference on Artificial Intelligence in Amsterdam, the Netherlands, 1994.

[2] M. Davis, "Media Streams: Representing Video for Retrieval and Repurposing," PhD dissertation, Cambridge, Massachusetts: MIT Media Laboratory, 1995.

[3] R. W. Picard, "Light-years from Lena: Video and Image Libraries of the Future," International Conference on Image Processing, Washington DC, October, 1995.

[4] A. Pentland, R. W. Picard, S. Sclaroff, "Photobook: Content-based Manipulation of Image Databases," *SPIE Storage and Retrieval Image and Video Databases II*, No. 2185, February 1994.

[5] R. W. Picard, "A Society of Models for Video and Image Libraries," MIT Media Laboratory Perceptual Computing Section Technical Report No. 360, April 1996.

[6] R. W. Picard, T. P. Minka, "Vision Texture for Annotation," International Conference on Image Processing, Washington DC, October 1995.

[7] W. Niblack, R. Barber, W. Equitz, M. Flickner, et al., "The QBIC project: Querying images by content using color, texture and shape," *Storage and Retrieval for Image and Video Databases (W. Niblack, ed.)*, PP. 173-181, SPIE, February 1993.

[8] S. Santini, R. Jain, "The Graphical Specification of Similarity Queries," http://vision.ucsd.edu/ ssantini/dbliterature.html.

[9] P. M. Kelly, M. Cannon, "Query by image example: the CANDID approach," http://www.c3.lanl.gov/ kelly/CANDID/pubs.shtml.

[10] C. Frankel, M. J. Swain, V. Athitsos, "WebSeer: An Image Search Engine for the World Wide Web," Technical Report 96-14, Computer Science Department, The University of Chicago, August 1996.

[11] G. Stix, "Finding Pictures on the Web," *Scientific American*, March 1997.

[12] T. Hou, A. Hsu, M. Chiu, "A Content-based Indexing Technique using relative Geometry Features," *SPIE Image Storage and Retrieval Systems*, Vol. 1662, 1992.

[13] H. Zhang, D. Zhong, "A Scheme for Visual Feature-based Image Indexing," *SPIE Image Storage and retrieval Systems*, Vol. 2420, 1995.

[14] R. L. Delanoy, R. J. Sasiela, "Machine Learning for A Toolkit for Image Mining," Lincoln Laboratory 1017, MIT, Lexington, MA, March 1995.

[15] S.K. Chang, "Extending Visual Languages for Multimedia," *IEEE Multimedia*, Vol. 3, No. 3, Fall 1996.

[16] J.K. Wu, A. D. Narasimhalu, B.M. Mehtre, C.P. Lam, Y.J. Gao, "CORE: A Content-based Retrieval Engine for Multimedia Information Systems," *Multimedia Systems*, 3:25-41, 1995.

[17] S. Lee, F. Hsu, "Spatial Reasoning and Similarity Retrieval of Images Using 2D C-String Knowledge Representation," *Pattern Recognition*, Vol. 25, No. 3, 1992.

[18] N. A. Van House, M. Butler, "User-Centered Iterative Design for Digital Libraries," *D-Lib Magazine*, February 1996.

[19] S. W. Smoliar, H. Zhang, "Content-based video Indexing and Retrieval," *IEEE Multimedia*, Summer 1994.

[20] J. Malik, P. Perona, "Preattentive texture discrimination with early vision mechanisms," *Journal of the Optical Society of America*, 1990.

[21] E. Kandel, "Perception of Motion, Depth and Form," *Principles of Neural Science*, 1991.

[22] S.K. Chang, "Principles of Pictorial Information Systems Design," Prentice-Hall, Englewood Cliff, NJ, 1989.

[23] E. Saber, A. Tekalp, et al., "Annotation of natural scenes using adaptive Color Segmentation," *ISET/SPIE Electronic Imaging*, February 1995.

[24] R. L. Delanoy, J.G. Verly, D. E. Dudgeon, "Machine Intelligent Automatic Recognition of Critical Mobile Targets in Laser Radar Imagery," Lincoln Laboratory, J.6, 161-186, Spring 1993.

[25] J. R. Smith, S.F. Chang, "Searching for Images and Videos on the World-Wide Web," Technical Report No. 459-96-25, Center for Telecommunication Research, Columbia University, August 1996.

[26] W. Horton, "The Icon Books: Visual Symbols for Computer Systems and Documentation," John Wilkey & Sons, Inc, 1994.

[27] C.K. Bliss, "Semantography-Blissymbolics," 3rd ed., Sydney, N.S.W., Australia, Semantography-Blissymbolics Publications, 1978.

[28] G. Pass, R. Zabih, J. Miller, "Comparing Images Using Color Coherence Vectors," Proceedings of the Fourth ACM International Multimedia Conference, Boston, November, 1996.

[29] V. Ogle, M. Stonebraker, "Chabot: Retrieval from A Relational database of Images," IEEE Computer, 28(9):40-48, September 1995.

[30] M. Minsky, The Society of Mind, Simon & Schuster, Inc, 1986.

[31] W. Horton, The Icon Book: Visual Symbols for Computer Systems and Documentation, John Wiley & Sons, Inc., 1994.

[32] O. Neurath, International Picture Language, New York: State Mutual Book and Periodical Service, 1981.

[33] Network Wizards, http://www.nw.com/zone/WWW/report.html, 1996.

[34] WebCrawler, http://webcrawler.com/WebCrawler/Facts/Size.html, 1996.

[35] H. Dreyfuss, Symbol Sourcebook: An authoritative Guide to International Graphic Symbols, New York: McGraw-Hill, 1972.

[36] M. Davis, "Media Streams: An Iconic Visual Language for Video Representation," in: Readings in Human-Computer Interaction: Toward the Year 2000, ed. Ronald M. Baecker, Jonathan Grudin, William A. S. Buxton, and Saul Greenberg. 854-866. 2nd ed., San Francisco: Morgan Kaufmann Publishers, Inc., http://www.interval.com/papers/mediastreams, 1995.

[37] R.E. Griswold and M.T. Griswold, The Implementation of the Icon Programming Language, Princeton University Press, Princeton, New Jersey, 1986.

[38] K. Haase and W. Sack, "Framer Manual," internal document, Cambridge, Massachusetts: MIT Media Laboratory, 1993.

[39] W.H. Huggins and D.R. Entwisle, Iconic Communication: an annotated bibliography, The Johns Hopkins University Press, Baltimore, Maryland, 1974.

[40] M. McLuhan, The Getenberg Galaxy: the Making of Typographic Man, Toronto: University of Toronto Press, 1962.

[41] O. Neurath, International Picture Language. New York: State Mutual Book and Periodical Service, 1981.

[42] S.L. Sieglinde, "The Web: How Big? How Far?" European University Institute, Florence, Italy,
http://www.cali.org/96conf/schreine.html, 1996.

[43] S. Lee, F. Hsu, "2D C-string: A New Spatial Knowledge Representation for Image Database System," Pattern Recognition, Vol. 23, No. 10, pp. 1077-1087, 1990.

[44] G. Furnas and J. Zacks, "Multitree: enriching and reusing hierarchical structure," Human Factors in Computing Systems CHI'94 Conference Proceedings, Boston, MA, ACM, pp. 330-336, 1994.

[45] G. Furnas, T.K. Landauer, et al., "The vocabulary problem in human-system communication," Communications of the Association for Computing Machinery, 30(11), pp. 964-971, Nov 1987.