An Analytical Study of Browsing Strategies in a Content-Based Image Retrieval System

Alison Gilchrest
Holley Long

Prepared as a term project for Dr. Gary Marchionini
INLS 180-01
Spring 2001

Introduction

In Fall 1999, the journal Library Trends devoted an entire issue to the topic "Progress in Visual Information Access and Retrieval." Two years prior, professor Edie M. Rasmussen of University of Pittsburgh published a comprehensive literature review titled "Indexing Images" in the Annual Review of Information Science and Technology. As these works (and many other more recent publications attest), there continues to be a growing body of literature devoted to the scholarly issues and practical concerns of storage, indexing, and retrieval in image-based information systems. Little of this work, however, addresses how these systems are used by end users, what types of interfaces afford better retrieval, and what theoretical approaches to image classification translate best to different types of image systems.

The opportunity to investigate a personal interest for this end of term project encouraged both of us to pursue an inquiry related to these questions. We both come to SILS with undergraduate degrees in History of Art, and we both have independently fostered an interest in image retrieval – specifically developments in content-based image retrieval (CBIR) systems – during our time here. This project then, represented a fortunate opportunity for us to collaborate and grow in our learning.

When considering information seeking behavior, we looked back at the literature covered throughout the semester in this course and settled upon the unanswered questions in Barbara Kwasnik's article as a point of departure. Specifically, we looked at the following research questions: Are Kwasnik's functional components of browsing applicable to a CBIR environment? What strategies do users employ to find images in a content-based image retrieval system, and does the nature of the query affect the strategy?

In conjunction with these questions, we have applied the subsequent operational definitions to further define our analysis. First, the term "strategies" is used interchangeably throughout this paper for the phrase "functional components." While we acknowledge there are semantic differences between the two, for the purposes of this study, we are using them as equivalent terms. Secondly, "users" encompasses two distinct populations – ourselves as subjects of analysis, and users-at-large as recorded in system transaction logs. Finally, "nature of the query" is a concept that is explored and defined at length in section B below.

II. Background and Literature Review

A. The Kwasnik Framework

In her article, "A Descriptive Study of the Functional Components of Browsing," Barbara H. Kwasnik investigates the various strategies involved in browsing and what roles they serve. Her study involved observing thirty participants browsing through a Columbia House Record Club catalog in three formats.

In her literature review Kwasnik establishes the definition of browsing as "the strategic and adaptive technique that people use to search, scan, navigate through, skim, sample, and explore information systems"[1] or a "heuristic search in a well-connected space of records."[2] She also discusses factors that affect browsing, such as structured versus unstructured browsing and purposeful versus non-purposeful browsing. She points out that people will seek structure and purpose even where it seemingly does not exist.

Kwasnik[3] identified six key behaviors associated with browsing that guided her study. The first, *orientation*, involves learning the physical structure of the search environment, whether that structure is physical or conceptual. The process of orientation does not occur all at once, but gradually over the course of interaction with the system. The second strategy, *placemarking*, holds elements for future consideration. Placemarks can be held in a number of ways – for example, folding down the corner of a page or writing down a URL are both ways to indicate a location to which one might want to return. *Identification*, the third strategy, occurs when the user recognizes a potentially useful view or eliminates an unproductive path.

While browsing a user may encounter elements that are confusing or unclear. The *resolution of anomalies* is a natural part of browsing, and the fourth strategy. Kwasnik notes that users will resolve anomalies even if the element in question is not otherwise interesting. *Transitions* are movements from one view to another, where a view is defined as a person's "span of attention." A user can transition toward a potentially useful view or away from an uninteresting one. Finally, *comparisons* are a basic component of browsing that takes place at all

---

[1] Kwasnik, Barbara "A Descriptive Study of the Functional Components of Browsing," Engineering for Human-Computer Interaction (1992): 191.
[2] *Ibid*, 193.
[3] *Ibid*, 195.

levels. A user may compare one element to another, parts of the structure to other parts of the structure, or the whole environment to other environments.[4]

Although the article was published before the study concluded, and no further work was carried out, Kwasnik raises several interesting questions:[5]

- How does choosing a starting and ending point affect browsing strategies?
- How does the environment's structure affect browsing? How does it affect movement? How does the structure influence orientation?
- How do people reorient themselves when they are lost?
- Do transitions away necessitate different navigation tools than transitions toward an item?

As many of these questions remain under-studied, we have selected Kwasnik's functional components of browsing as the framework for this analysis. By revisiting the six categories that she identified and applying them to the context of an image-only content-based image retrieval system, we aim to address the research question of whether or not these strategies translate to such an environment, and if so, in what ways are they applicable?

B. The Nature of Image Queries

Due to improvements in image retrieval technology and the proliferation of picture databases, scholars have begun to study the way people form image queries and how they search for visual materials. In particular, a good deal of research focuses on visual image query classification and categorization. The following discussion reviews four important studies conducted in this area.

Peter Enser was one of the first IS scholars to study the problem of image classification. In 1997 Enser and Linda Armitage expanded on a preliminary Enser study to design a faceted schema for image classification.[6] They studied seven libraries, each containing picture archives. Together these collections represented a wide range of topics, varying levels of specialization, and a broad clientele base.

After analyzing 200 queries from each library, Armitage and Enser developed a schema based on the work of art historian Irwin Panofsky. In their approach, images are divided into the iconographical, pre-iconographical, and iconological. Pre-iconography queries are satisfied with

---

[4] Kwasnik, 195.
[5] *Ibid,* 196.
[6] Armitage, Linda and Peter Enser, "Analysis of User Need in Image Archives," Journal of Information Science 23, no. 4 (1997): 287 – 299.

generic images. For example, a picture of "an apple" falls into this category. Iconographical queries require specific instances of a person place or thing. "A photograph of the apple dropping in Times Square, New Years Eve 1989" would be considered an iconographical image. Iconological images represent abstract ideas or emotions, an example of which might be a wood cut of Eve holding an apple to symbolize original sin. These three categories are further subdivided by the descriptors who, what, where, and when, where each of these facets is assigned a code. For the aforementioned example, "a photograph of the apple dropping in Times Square, New Years Eve 1989," the code is S2 (individually named event) + S3 (individually named geographical location) + S4 (linear time). The code for a wood cut of Eve holding an apple to symbolize original sin is A1 (mythical being) + G1 (kind of thing) + A2 (emotion or abstraction). The following table represents the Armitage and Enser approach to image classification:

	ICONOGRAPHY (Specifics)	**PRE-ICONOGRAPHY** (Generics)	**ICONOLOGY** (Abstracts)
WHO?	Individually named person, group (S1)	Kind of person or thing (G1)	Mythical or fictitious being (A1)
WHAT?	Individually named event, action (S2)	Kind of event, action, condition (G2)	Emotion or abstraction (A2)
WHERE?	Individually named geographical location (S3)	Kind of place: geographical, architectural (G3)	Place symbolized (A3)
WHEN?	Linear time: date or period (S4)	Cyclical time: season, time of day (G4)	Emotion, abstraction, symbolized by time (A4)

Figure 1 [7]

Armitage and Enser conclude their discussion by suggesting that their schema could be applied to the interface of an image databases to improve retrieval. It would be difficult however, to apply this schema to a CBIR system. How could iconological subjects be derived from features like color and texture? Furthermore, the who, what, where and when facets are best described with metadata.

In the book chapter "User Types and Queries: Impact on Image Access Systems," Lucinda Keister[8] examined queries submitted to the National Library of Medicine's Prints and

[7] Armitage and Enser, 290.

Photographs collection and discovered that queries must be considered in the context of the use and the user. Keister reconstructed queries submitted to NLM over a one-year period from notes taken by the library staff. She focused on who submitted the query and how they formed the search question. Keister estimated that one half of the library's clientele were publishing professionals, one third were health professionals, and the rest came from museum and academic communities and the general public.[9] Her analysis suggested that patrons do not request visual information in a consistent manner. For example, imaging and photographic professionals frame their queries in a visual way: "an action shot of George Papanicolau, has to be horizontal and in color." Health professionals asked subject oriented queries: "Do you have pictures of cholera?" Patrons from museums and academia often had exact citations: "'The Cow-Pock, or the Wonderful Effects of the New Inoculation' by James Gillary."

Keister found that users' queries often contained subject terms, but more often they went beyond a basic topic request. Patrons often use words to build a visual construct for an image that they know exists or one that they imagine would satisfy an information need. She uses the example of Benjamin Rush's "Tranquilizing Chair" to illustrate this point, as rather than asking for the image by name, users were more inclined to request the picture of "the man sitting in the chair with a box on his head."[10] Lastly, Keister's study confirmed that people are seeking to use an image in a context different from its original intended purpose, which influences the terms they choose when describing their query.

The following quote concisely summarizes Keister's research: "It is not so much that a picture is worth a thousand words, for many fewer words can describe a still picture for most retrieval purposes. The issue has more to do with the fact that those words may vary from one person to another."[11]

Corinne Jörgensen took a different approach to the problem of image classification.[12] She showed a group of 50 participants six images selected at random and asked them to describe the images based on a particular task. The first group performed a descriptive viewing task in which they were asked to describe what they noticed about the image. The second group

---

[8] Lucinda H. Keister, "User Types and Queries: Impact on Image Access Systems," Challenges in Indexing Electronic Text and Images, (1994): 7 – 22.
[9] *Ibid*, 7 – 8.
[10] *Ibid*, 13.
[11] *Ibid*, 17.
[12] Jörgensen, Corrine, "Testing an Image Description Template." Proceedings of the 59th Annual Meeting of the American Society for Information Science, Baltimore, MD. (Oct 1996): 21-24: 209 – 213.

performed a descriptive search task in which they were asked to describe the image as if they hoped to find it in a collection of images. The last group (descriptive memory task) wrote their descriptions four weeks after viewing the images, based on what they could remember. From this data, she identified three major categories of image attributes: perceptual, interpretive, and reactive. Perceptual attributes are the result of direct visual stimulus and include categories like "color" or "object." Interpretive attributes like "style" require the user to apply a general level of knowledge or make some inference from the basic visual clues. Reactive attributes describe personal reactions to an image. Within these broad categories are 12 classes, described below:

PERCEPTUAL	
LITERAL OBJECT	Literal (visually perceived) objects
PEOPLE	Presence of a human form
COLOR	Specific named colors and terms relating to color like value, hue, tint
LOCATION	General and specific location of picture components
VISUAL ELEMENTS	Terms for which there is a body of empirical neurophysiological evidence, such as orientation, shape, visual component, or texture
DESCRIPTION	Descriptive adjectives and words referring to size or quantity
INTERPRETIVE	
PEOPLE QUALITIES	Qualities such as the nature of the relationship among people depicted in an image, their mental or emotional state, or occupation
ART HISTORICAL INFORMATION	Information which is related to the production context of the image, such as artist, medium, style, and type
ABSTRACT CONCEPTS	Abstract, thematic, and symbolic image descriptors
CONTENT / STORY	Attributes relating to a specific instance being depicted, such as activity, event, and setting
REACTIVE	
PERSONAL REACTION	Personal reaction to the image
EXTERNAL RELATIONSHIP	Comparison of attributes within a picture or among pictures or reference to an external entity

Figure 2 [13]

Once Jörgensen created these categories, she calculated the distribution of each class by task and concluded that participants assigned attributes based on the given task. A task that required image description elicited perceptual attributes, while searching or sorting tasks elicited interpretive and reactive attributes. Her work suggests that it is not enough to look at the image by itself; rather, classification schemes should be designed with the retrieval task in mind.

Raya Fidel's research further supported Jörgensen's arguments.[14] Fidel took one hundred queries from a stock photo agency and applied Jörgensen's attribute classes to classify the queries. Her findings indicate that the way an image will be used affects how a person searches

---

[13] Jörgensen, 210.
[14] Fidel, Raya. "The Image Retrieval Task Implications for the Design and Evaluation of Image Databases," The New Hypermedia and Multimedia, (1997): 181 – 199.

for that image. According to Fidel's findings, images can either be sources of information or objects to serve some other purpose. For example, a painting of a 19th century English interior could be studied by a social historian for the information it contains about domestic life in the Victorian era, but the same painting could be used as cover art for a Victorian mystery novel. Requesting the image for the former use is an illustration of a "data pole" query. (Other examples might include maps, x-rays, or blueprints.) Requesting the image to use for a book cover is an instance of a query on the "object pole," where the image will be used to construct or enhance some other object. In Fidel's model, images that adorn are on the objects pole. She does note however, that most image queries do not fall neatly into one category or the other; rather, most fall on a continuum between the two poles.

Fidel points out that queries from the data and object poles are two entirely different retrieval tasks that require contrasting search strategies. Below is a table from "The Image Retrieval Task: Implications for the Design and Evaluation of Image Databases" which summarizes the differences between data pole and object pole searching behaviors:

DATA POLE	OBJECT POLE
Images provide information	Images are objects
Relevance criteria can be determined ahead of time	Users will recognize relevance criteria 'when they see them'
Relevance criteria are specifications of which the user is aware	Relevance criteria are latent and are invoked when viewing images
It is possible for users to explain why an image is relevant	It might be difficult for users to explain why an image is relevant
Images can be retrieved with textual and other verbal clues	It might be difficult to find verbal clues for retrieval, clues are often visual
Color, shape, and texture can convey information and therefore are important for retrieval	No evidence exists that color, shape, and texture are important for retrieval
Images must include similar information to satisfy the same need	Two very different images may satisfy the same need
Of-ness often equals aboutness	Of-ness is likely to be different from aboutness
Biographical attributes are not likely to play a role	Biographical attributes are important for relevance assessment
To satisfy requests may require sets of more than one image	Requests are usually satisfied with one image
May not require browsing through the whole answer set	Requires browsing through the whole answer set
Browsing is time consuming	Browsing can be done rapidly

Figure 3 [15]

Many of these points are important when considering search strategies in content-based image retrieval systems. Queries on the data pole are more easily described with words,

[15] Fidel, 191.

enabling a searcher to identify ahead of time precisely what would satisfy his information need. Object pole searching however, is very difficult to do verbally. Searchers in this category are more likely to say, "I'll recognize what I'm looking for when I see it."

Data pole searching does not require a complete review of the retrieved set. Once a searcher finds an image that supplies the necessary information, he or she can stop looking. Searchers who want an image on the object pole often browse through the entire retrieved set before selecting the most relevant picture. Furthermore, browsing takes less time and effort for object pole queries than data pole queries.

Content-based image retrieval literature often categorizes image queries in four broad classes: content-independent, content, concept, and abstraction. Content-independent queries search on attributes that exist outside of the image itself like "date of creation" or "title." As these elements cannot be identified by a CBIR system, the only way to allow querying on this information is to include human generated metadata. Content-dependent attributes are visually perceivable elements like color, texture, and shape, which current technology is adept at identifying, extracting, and indexing. Concept attributes are semantic elements of a picture that require logical inferences about the identity of objects in the image. Haering et al[16] have developed a CBIR system that identifies trees (content attribute) based on the shapes of the leaves (concept attribute). Abstract attributes, like emotions, are difficult for CBIR systems to recognize. Identifying abstract attributes necessitates high-level reasoning about the meaning and purpose of the objects or scenes depicted and often draws on culture-specific knowledge.

Adapting this CBIR schema provides a framework to merge the previously described research:

[16] N. Haering et al. "Locating deciduous trees" in Proceedings of IEEE Workshop on Content-Based Access of Image and Video Libraries, San Juan, Puerto Rico (June 1997): 18-25.

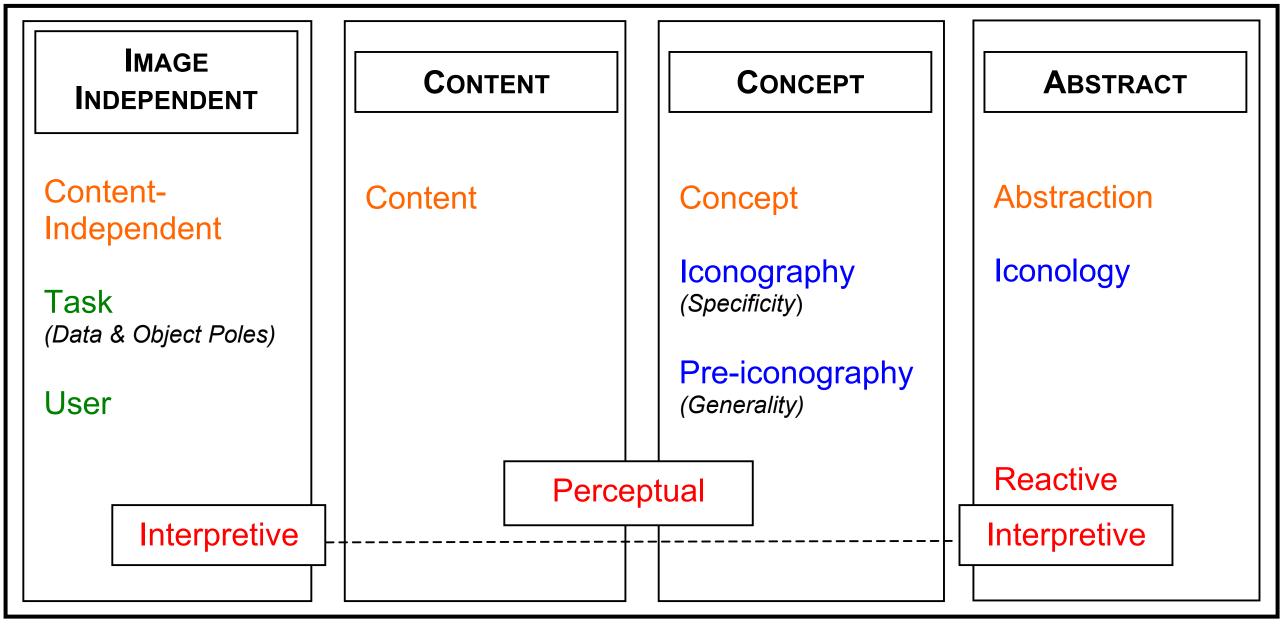


Figure 4: Merged query schemas

Enser and Armitage's Pre-Iconography and Iconography attributes are conceptual and differ only in their level of specificity. Iconology attributes require higher-level reasoning and culture-specific knowledge and therefore belong in the abstract category. Those classes identified by Jörgensen do not fall neatly into single categories in the above schema. Rather, her "perceptual" class includes content attributes (e.g. color) and concept attributes (e.g. the name of the literal object) while "interpretive" attributes call for the application of outside knowledge. This could mean recognizing the meaning behind a symbol (abstract) or identifying an art historical period (content-independent). Jörgensen's research also suggests that there is a relationship between the nature of the task and image classification. Fidel extended this idea to include data and object poles. She suggests that images are used as information sources and as articles in a construction; therefore the way the image will be used affects how it is retrieved. Finally, Keister points out that the user's context influences how a query is formulated. A publisher may ask for a photograph with a vertical orientation that illustrates the effects of poor sanitation on health. Museum professionals and historians often have exact citations, while a health professional may simply ask for pictures of cholera. These attributes – external to the image itself – are content independent.

C. Content Based Image Retrieval: SIMPLIcity

The proliferation of image content on the web has stimulated enormous growth in the research community as to how to handle search and retrieval of large image databases. Image search engines continue to be developed in both the commercial and academic sectors, each one incorporating varying degrees of content and concept-based retrieval techniques and algorithms. Content-based image retrieval systems are defined by automatic indexing of features such as color, texture, shape, and spatial relation. In this study, we conducted our investigation using the SIMPLIcity system, a web-based experimental CBIR hosted by Professor James Wang of Penn State University.

Examples of well-known commercial CBIR systems include IBM's Query By Image Content (QBIC),[17] which currently provides two demonstration sites and has also been incorporated into the Hermitage museum's website to allow visitors to search the collection. The search engine AltaVista incorporates the VIR Image engine system produced by VIRAGE, Inc.,[18] and as of May 2000, Yahoo partnered with Interpix Software[19] to provide image searching capability for both still and video images via their search portal.

Academic demonstration systems have also gained a lot of attention for their advances in CBIR. Columbia's WebSEEK[20] system allows users to search a combined database of photographs and video clips by using broad categories of search terms, then refining the query through color selection, histogram adjustment, and relevance feedback. Available as Unix/Linux freeware, MIT's Photobook[21] system incorporates shape and texture extraction algorithms that have shown particular success in retrieving human faces. The package also provides a wide variety of feature types to select from, enabling users to select the most appropriate features for a particular query. NeTRA2,[22] developed at UC Santa Barbara, uses an image segmentation and regional color-matching algorithm that allows users to select regions of interest as queries. UC Berkeley's BlobWorld[23] system also uses a region-based approach that allows users to rank the importance of features within image segments based on characteristics such as color, texture,

---

[17] http://wwwqbic.almaden.ibm.com/
[18] http://www.virage.com/customers/success_alta_vista.html
[19] http://docs.yahoo.com/docs/pr/release21.html
[20] http://disney.ctr.columbia.edu/webseek/
[21] http://www-white.media.mit.edu/~tpminka/photobook/
[22] http://maya.ece.ucsb.edu/Netra/index2.html
[23] http://elib.cs.berkeley.edu/photos/blobworld/

location, shape, and size. Both NeTRA2 and BlobWorld have similar backend image databases of Corel PhotoCD stock photography.

The system chosen for our analysis is an academic demonstration system called SIMPLIcity: Semantics-sensitive Integrated Matching for Picture LIbraries. Developed by Professors James Z. Wang and Gio Wiederhold of Stanford University as part of the NSF Digital Library Initiative II (DLI-II) program, it builds on feature and region-based systems, but also incorporates what they term "semantics-sensitive" classification. "The system classifies images into semantic categories, such as textured-nontextured, graph-photograph. Potentially the categorization enhances retrieval by permitting semantically-adaptive searching methods and narrowing down the searching range in a database."[24] Wang relocated from Stanford to Penn State in 2000, taking SIMPLIcity with him. It continues to undergo development and monitoring, including monthly use statistics gathering.[25]

Several factors influenced the decision to use SIMPLIcity for our study. Like NeTRA and BlobWorld, it utilizes a backend database of over 200,000 Corel PhotoCD stock photography images. This consistency makes our results potentially generalizable or applicable to future studies that compare cross-system features or performance. The tremendous quantity and variety of images available enabled us to formulate queries based upon a number of different image classification schemas with relative certainty that images exist to satisfy the queries.

Of all of the academic demonstration systems we looked at, SIMPLIcity had other key advantages: first, it offers the most images per screen for a retrieved set (32 per page, compared with, for example, 6 for NeTra2 and 20 for BlobWorld). Secondly, SIMPLIcity had the most search tools available, providing the greatest variety of inroads into the system. It allows users to being by selecting a random set of images, draw their query using a java-based drawing tool, or start with one of four preselected query images. Users can also enter an image number directly if they know what image they are looking for (or want to guess at random), or they can enter a URL for any image pointer on the Internet to use as a query.

[24] James Z. Wang, Jia Li, Gio Wiederhold, ``SIMPLIcity: Semantics-sensitive Integrated Matching for Picture LIbraries," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, 2001.
[25] http://wang.ist.psu.edu/docs/log/

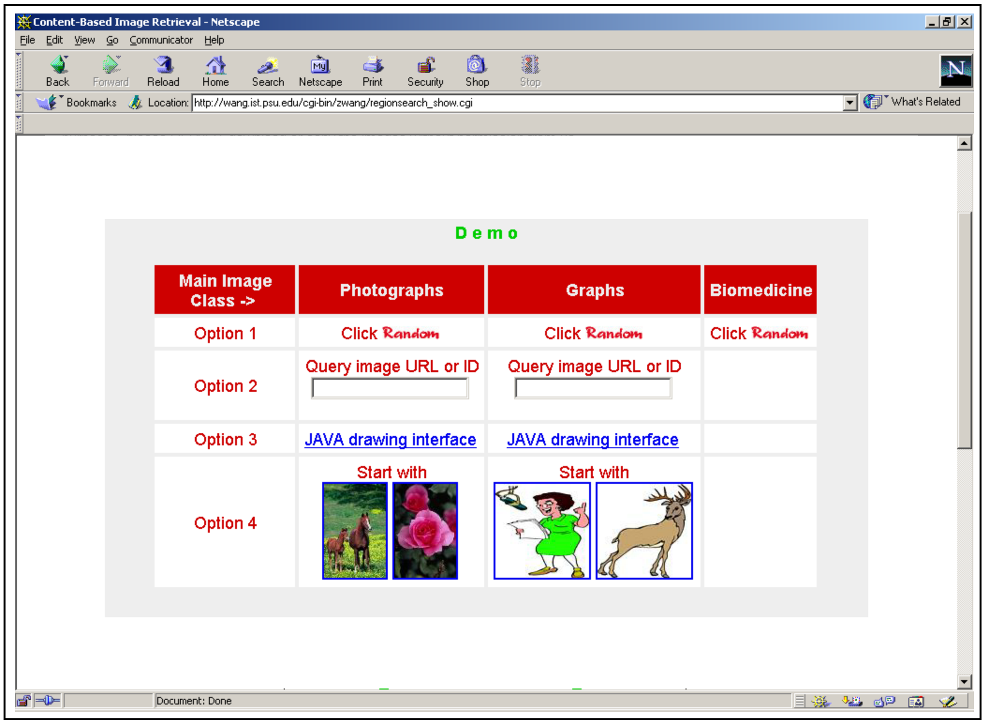


Figure 5: Query interface for the SIMPLIcity Demo system

By evaluating both the logs and our own query tasks, we were able to determine which of these tools were most often employed in this content-based retrieval environment, and observe what types of strategies are most suited to particular classes of image queries.

The third key advantage to using SIMPLIcity is that, unlike most other systems, it does not provide any means of keyword or concept searching – thus the system effectively limits users to visual expressions of an information need. Among the many pitfalls this avoids are the often-cited criticisms that keyword indexing is both insufficient to capture the true "meaning" of an image, and inter-indexer consistency for image description is very low. Additionally, when query terms fail to match the index term and recall is poor, it raises the question of using a controlled vocabulary for both indexing and query formulation – a measure that is time consuming, costly, and difficult for novice users to understand. Additionally, as most CBIR systems are built for the express purpose of avoiding concept-based retrieval, it simply made sense to use a system with no concept-based attributes.

The final factor that influenced our decision to use SIMPLIcity for this study was the generous and polite cooperation we received from Professor Wang. He responded to our inquiries promptly, expressed interest in our work, and provided us with log files from the web server for analysis.

I. Log Analysis

A. Methodology

Our initial intent with web log analysis was to be able to reconstruct full user sessions as well as gather usage statistics for the SIMPLIcity system. As it happens, we were unable to determine enough session data that would enable us to recreate 3rd party searches with any degree of confidence. Instead, we focused on extracting the general usage data that follows.

Upon receiving the logs for the month of March 2001, we first attempted to parse each type of request and assign it a meaning that corresponded to a user action on the system. When this proved too ambiguous, we developed a test task – a methodical series of searches designed to utilize every aspect of the system. We performed these tasks while taking screen shots corresponding to each step, and then requested the logs that recorded this session. Due to some unexplained difficulty however, the log for our session stopped recording after task 14, where we entered the java drawing interface.[26] However, using our task list and screen shot record we were still able to determine nearly all of the other calls with a high degree of certainty and construct a key by which to analyze the log.

For the month of March, the SIMPLIcity web server received 222,396 total requests from 127,510 unique IP addresses.[27] Due to the sheer volume of traffic on this site, we decided to limit our analysis to the first 14 days of the month, or approximately half of the total log size. All entries from ils.unc.edu and psu.edu were struck, as well as those from Wang's partner's site at stanford.edu under the assumption that these sessions do not represent typical searching behavior. In addition, one other anomalous session was struck: an unregistered IP address that rapidly called all images in sequential order for a period of more than 10 hours, assumed to be a spider. We then further eliminated repeat sessions from duplicate IP addresses to isolate searching and browsing behaviors by all users who were (assumed) new to the system.

B. Analysis

By using our log key, we gathered use statistics based upon how these users entered the system. We also chose to look at how many users viewed single enlarged images as a potential

[26] We have identified some possible factors that could have contributed to this error – Dr. Wang had indicated to us via email that he was experiencing server troubles and denial of service attacks during the time in which we were performing our study; also, the java interface is an application housed on different web server, and thus we concluded the logs did not record these requests.
[27] http://wang.ist.psu.edu/docs/log/www2001/totals0301.html

measure of query satisfaction. Other measures we would have liked to have extracted from the logs had time permitted include the number of repeat visitors, how many times they returned, and the entry points chosen on subsequent visits. We are also interested in calculating an average number of retrieved sets per session to give an indication of how long a user stays in the system apart from clock time.

Our analysis shows that 95.5% of users started with the search demo interface shown in figure 6. Those who did not had an entry point that was either a call for a random set, or it was indeterminate from the log. The random feature had the highest percentage of first-click entry points at 25.4%, with 6282 (the horses) and 27871 (the roses) nearly equal at 16.4% and 14.9% respectively. The clip-art image of the deer, 125007, had the lowest percentage of use at 1.5%.

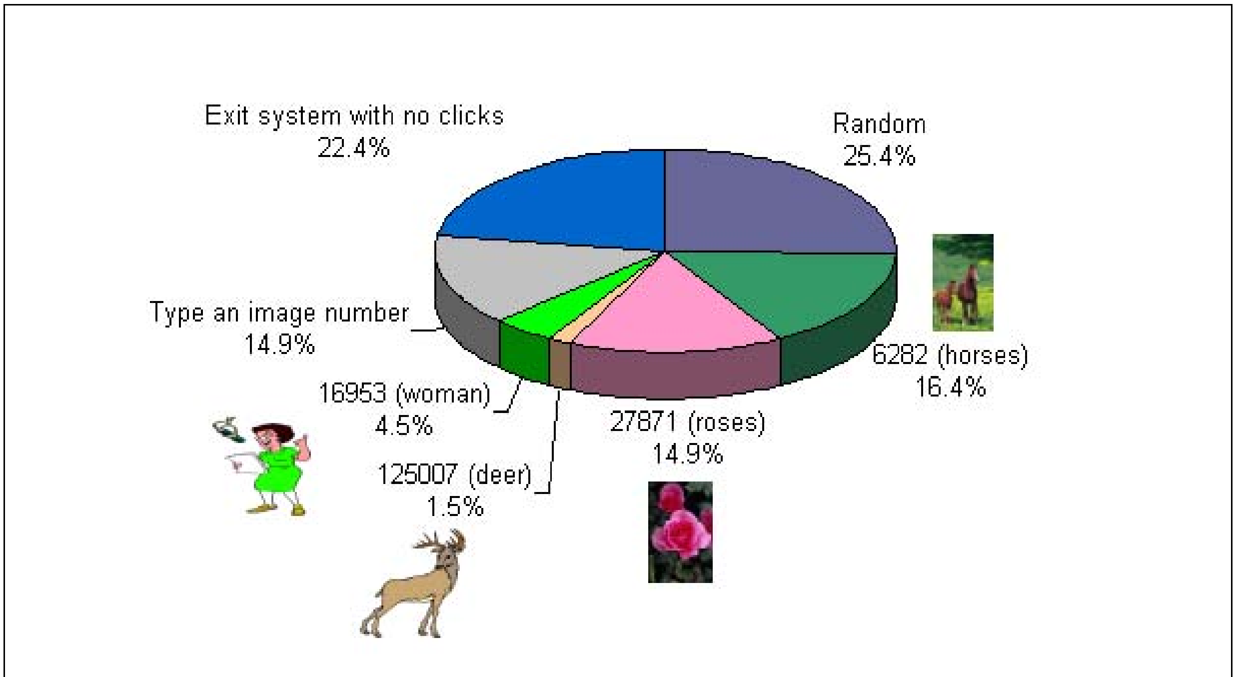


Figure 6: Summary of first-click entry points from web logs

Fully 22.4% of users are recorded as exiting the system without executing any searches – either that, or the logs failed to record any calls beyond the initial loading of the search demo interface. As we were unable to identify from this server's logs when the java drawing interface was selected, it is possible that some of these truncated sessions were actually drawing interface calls that "dropped out" just as our test session did.

Of the 14.9% of users who began searching by entering a number directly into the form, 60% (or about 9% of the total analyzed log entries) were recorded as the number 0. This struck us as anomalous because most file numbering schemes intuitively start at 1, and the odds of that many users entering 0 as a first query seem very low, especially in light of all of the other available entry points. Though SIMPLIcity will return an image with the file number 0 and an accompanying retrieved set if one enters “0” into the form, further investigation showed that 0 is also the default returned when a user enters a URL into the form without the http:// protocol prefix.[28] This may help explain the large number of occurrences of this particular entry point in the logs.

Finally, we looked at the number of sessions in which users selected a thumbnail image for enlarging. A total of 11.9% of the analyzed sessions included at least one call in the log to a single enlarged image. The mean number of enlarged images per session was 2.25, the median was 1.5, and the mode was 1. The maximum number called during one session was five. A full 38.8% of enlarged images depicted female models, while the rest consisted primarily of animals, flowers, or buildings. From this particular facet of the log analysis, we are unable to make any conclusive judgments as to whether or not these images satisfied a query or information need, or if they simply represent users exploring within the system. From our own experience with the system interface, we see it is quite possible that enlargements represent users attempting to resolve anomalies, as often times the thumbnails are too small to see image details.

IV. Task Analysis

A. Methodology

To test the applicability of Kwasnik’s functional components of browsing in a CBIR system, we observed and recorded each other’s behavior while performing tasks in the SIMPLIcity image database.

Based on the image classification literature, together we developed nineteen query tasks[29] (see Appendix A for a complete list of queries). The first three were specific instances of images located in the database: two of which were images we selected for each other, the third was an image selected by a third party for our inter-rater reliability task. The next four tasks corresponded to Enser’s Pre-Iconography and Iconography in the concept category. Four queries from the abstract category contained iconological and affective attributes. Two tasks included both content and abstract aspects. To represent image independent queries, we used Fidel’s

[28] Form entries that are URLs for pages and not images either display an error message or a random image.
[29] See Appendix A for a complete list of queries.

object/data poles concept in four queries. The last two queries contained strong content elements, but also drew upon the idea of the user's task and context.

We then created a data collection instrument[30] based on Kwasnik's functional components of browsing. The instrument recorded the query task, total search time, number of clicks, actions taken by the subject, placemarkers and identifications, transitions, anomalies, and comparisons. The first column recorded the click number and the second column tracked the users action – for example, "click random" or "click the back button." The Placemarkers and Identification column identified images that the subject believed might lead to the desired image. When the subject hit a "dead end" or felt lost, this was noted in the Orientation column. Transitions toward, defined as clicks on an image, and transitions away, defined as beginning at a new entry point, were registered in the Transitions column. Comments columns for anomalies and comparisons existed to log observations made by the subject. All references to images were recorded by the image's file name, a unique numerical identifier.

As indicated above, we first had a third party selected a random image in the system for our inter-rater reliability task. This image was opened on a blank screen with no identifying numbers or URL's. We each entered SIMPLIcity independently, performed the task, and completed our own data collection instruments. Side-by-side evaluation[31] showed our behavior to be quite similar, and we were thus satisfied that our results would be comparable.

Throughout the remainder of the query analysis, we alternated turns performing tasks and logging observations. The subject talked aloud as she searched indicating placemarkers, anomalies, and comparisons while the recorder noted these things on the data collection instrument. The subject was also instructed to describe her mental model of the query image at the beginning of the search and to indicate if her mental model changed as the search progressed. Illustrations, or "anatomies" of task 4, the general query, are included in Appendix D (and also corresponds to the sample data sheet in the previous appendix.)

B. Analysis

Results of the query task suggest that Kwasnik's components of browsing accurately reflect users searching behaviors in CBIR systems. In fact, comparison, transition, and placemarking proved to be essential tools for navigation and retrieval in SIMPLIcity.

---

[30] See Appendix B for a sample data collection instrument
[31] See Appendix C for anatomies of the inter-rater reliability task

Twenty-two percent of the comments recorded during the query tasks indicated comparative activities. As subjects, we both made comparisons between images within a retrieved set, between the query image and the retrieved set, and between searches as a whole. For queries 1 and 2, we toggled back and forth between the drawing tool and the query image to evaluate color selections. While several comments reflected concept-oriented comparisons, we most often noted differences in the content attributes of shape and color. There were more comparisons recorded in the earlier searches than in later ones. This fact and the nature of the comparative comments suggest that we were using a comparison of the query image with the retrieved set in order to learn how the system works. As we oriented ourselves in the environment, we made fewer of these comparisons. In this case one browsing component, comparison, supported another, orientation.

Transition, or movement from one view to another, is an integral part of searching in SIMPLIcity. Most actions taken by the user involve clicking on an image or initiating a new search and thus yield new views. In her article, Kwasnik asks, "Do movements away from something require different navigational aids than do movements towards something?"[32] In our experiment we concluded that transitions toward the desired image occurred when one of us clicked on a placemarker or other potentially useful image. Beginning at a new entry point signaled that we deemed the present view to be a dead end and were moving away from that view. It was difficult to determine whether a click on the back button was a movement toward or away from a view. In one sense we were moving toward a placemarker; on the other hand, we were moving away from a dead end. As figures 7 and 8 (below) show, movements toward images were more common than movements away. This might be explained in several ways. Using the back button to return to previous views often proved disorienting because the browser cache was inconsistent, or in some instances a mental placemarker was perceived to be one screen back, but in reality was several screens back. In many cases it was easier to click on a placemarker and move forward than it was to transition away. In this vein, Alison noted that her searching preference was to keep "moving on" instead of returning to previous retrieved sets.

[32] Kwasnik, 195.

Figure 7: Transitions Away

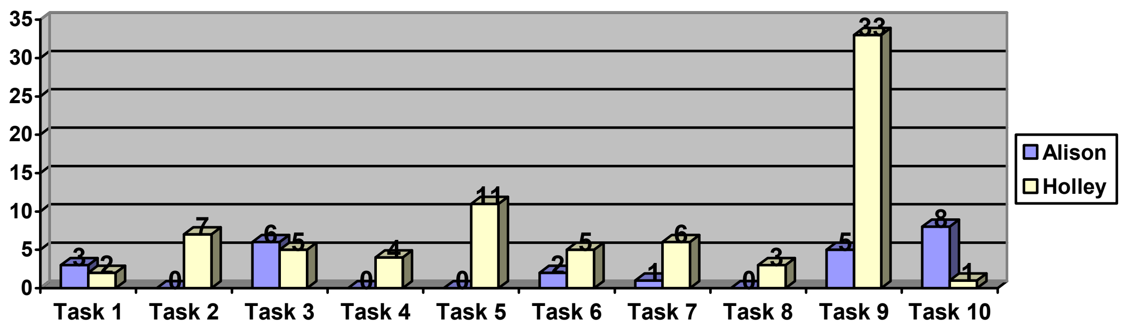
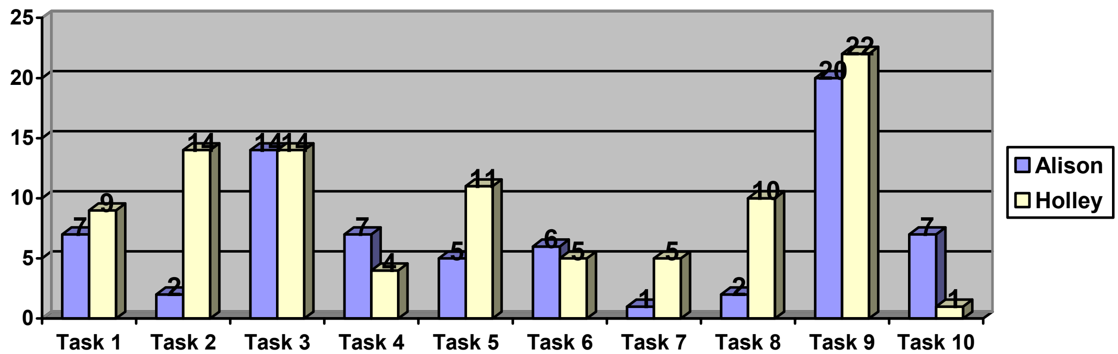


Figure 8: Transitions Toward



Number of placemarkers correlated to the time and number of clicks recorded for each task.[33] In most cases, we mentally placemarked images and returned to them by clicking the back button. We also relied on written placemarks and, on two occasions, typed them directly into the search box.

We both observed anomalies in both content and concept attributes. Early in the tasks, the anomalies observed were more about concepts. While looking for an elephant, Alison exclaimed with surprise when the system returned a set of dinosaur pictures. As the tasks progressed, we were more likely to notice incongruities in content attributes. The drawing tool seemed to be the most anomalous feature. When Alison drew yellow and green stripes in task 10, the system returned images with large regions of blue. After the drawing tool retrieved two anomalous sets, Alison abandoned that strategy and clicked random.

The other application of the resolution of anomalies component was in the case of thumbnails being too small to assess the content of the image with any certainty. Enlargement of

[33] See Appendix E for additional comparative graphs.

thumbnails then became a method by which we resolved uncertainty in order to established a relevance judgment.

When we designed the experiment, we expected identification to occur in conjunction with other components. We assumed that we would identify an interesting thumbnail and then placemark it. The identification of useless thumbnails would either be treated as anomalies or lead to a reorientation by starting at a new entry point. We found this, indeed, to be the case.

Our data and object query tasks confirmed many of the assertions Fidel makes about image searching behaviors.[34] It was much faster to browse for an object pole image to put on the front of a child's lunch box than it was to find a data pole image of a DNA double helix. It only took Alison two minutes and four clicks to find the lunch box image, while she gave up on the DNA double helix after seventeen minutes and twenty clicks. In task 8, Holley found several images would "do," but she searched several retrieved sets before selecting a picture for a camp brochure. This supports Fidel's observation that more than one image can satisfy an object pole query and that searchers seeking an object pole image will usually view the entire retrieved set before choosing a picture.

Contrary to Fidel's assertion, we found that searching on content features was more successful in object pole queries than data pole queries. For example, Alison's mental model of a bright and colorful image guided her quick selection of a lunchbox image, but when tasked with finding a DNA double helix she said, "I wish I knew what colors I am looking for. I'm confident I know what shape I'm looking for, but I'm not getting anywhere with that piece alone."

Logically, content-based queries should be best suited to content based image retrieval systems, while concept and abstract queries should be increasingly more difficult. Our tasks did not follow this pattern strictly. The general concept query tasks were completed in under six minutes with fewer than eleven clicks. Holley and Alison had vastly different results for the abstract and content query. While Holley found the content image (4 minutes) much faster than the abstract image (16 minutes), Alison found the content image in 16 minutes and the abstract image in 4 minutes. Alison's results may be attributed to a poorly chosen query task. Her content task: "Select an image for a university's publication that heavily features their school colors – yellow and green, and is of vertical orientation" contained elements of an object pole

[34] Fidel, 191.

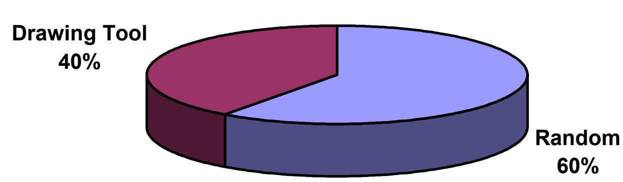
query. She found images that satisfied the “content” aspects of the query, but had difficulty meeting the data pole requirements.

The tremendous contrast between query task 3 (specific concept) and query task 4 (general concept) indicate that the level of specificity influences the outcome of a CBIR search. For both of us, the general concept query was the easiest to accomplish, while the specific concept took more time and clicks than any other type of query. The success of CBIR technology will depend on better access to concept specific images since many applications – newspaper archives, for example – often get requests for this category of query.

V. Discussion and Reflection

The inadequacies of the log analysis make it difficult to draw strong analogies between how users-at-large search SIMPLIcity and how we conducted our sample query tasks. We did learn however, that we limited ourselves to far fewer entry points than these users. (Compare Figure 6 above to Figure 9 here).

Figure 9: Our Task Entry Points



The most reasonable explanation for this dissimilarity is that we, as investigators, were conducting focused tasks and had progressively increasing familiarity with the system. Based upon the following findings: a) such a large percentage of visitors left without conducting searches at all; b) those that did query the system were about as likely to pick a “seed” image as they were random; and c) nearly 40% of the enlarged thumbnails were of female models, we feel it is safe to assume that most users, at least at this early stage, are simply playing in the system or are learning the features for the first time.

As stated above, had we had more time and more sophisticated means of analysis, we would have liked to have examined the log entries to extract features such as instances of repeat visitors, entry points on subsequent visits, and an overall average of number of retrieved sets per session.

Had we had more time and resources for the query analysis tasks, we would have preferred a more rigorous methodology that involved taking screen shots, or perhaps video/audio taping search sessions. We also would have enlisted subjects other than ourselves and had the procedures approved by the IRB. As for query selection, we have learned in retrospect that many of our sample queries are not the purest examples of each category. More thought to our choices would perhaps yield more valid, generalizable data.

In conclusion, we are pleased with the quality of our results and the amount of work we were able to accomplish on this project over the course of this semester.

Bibliography

Armitage, Linda H. and Peter G.B. Enser. "Analysis of User Need in Image Archives." *Journal of Information Science* 23, no. 4 (1997): 287 – 299.

Chen, Hsin-liang and Edie Rasmussen. "Intellectual Access to Images." *Library Trends*, Fall (1999): 291 – 302.

Fidel, Raya. "The Image Retrieval Task: Implications for the Design and Evaluation of Image Databases." *The New Hypermedia and Multimedia*, (1997): 181 – 199.

Haering, N. et al. "Locating Deciduous Trees." In *Proceedings of IEEE Workshop on Content-Based Access of Image and Video Libraries*, San Juan, Puerto Rico, (June 1997): 18 – 25.

Hastings, Samantha K. "Evaluation of Image Retrieval Systems: Role of User Feedback." *Library Trends*, (Fall 1999): 291 – 302.

Jörgensen, Corinne. "Access to Pictorial Material: A Review of Current Research and Future Prospects." *Computers and the Humanities* 33, (1999): 293 – 318.

--------. "Indexing Images: Testing an Image Description Template." *Proceedings of the 59th Annual Meeting of the American Society for Information Science, Baltimore, MD*. (Oct 1996): 21-24: 209 - 213.

Keister, Lucinda H. "User Types and Queries: Impact on Image Access Systems." In *Challenges in Indexing Electronic Text and Images*, ed. Raya Fidel et al., (1994): 7 – 22.

Kwasnik, Barbara H. "A Descriptive Study of the Functional Components of Browsing." In *Engineering for Human-Computer Interaction*, ed. J. Larson and C. Unger, 191 – 202. Amsterdam: Elsevier Science Publishers, 1992.

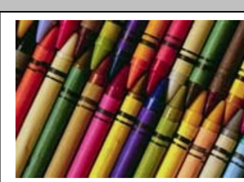
Rasumsen, Edie M., "Indexing Images." In *Annual Review of Information Science and Technology*, ed. M.E. Williams, 32: (1997): 169 – 196.

Wang, James Z. et al. "SIMPLIcity: Semantics-sensitive Integrated Matching for Picture Libraries." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23, (2001).

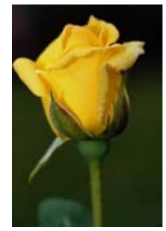
VI. Appendices

Appendix A: Queries

1. Inter-rater reliability task: specific instance: Image 47688

Holley's Query	Image that satisfied query	Alison's Query	Image that satisfied query
2. Specific Instance			
Image 5219		Image 59488	
3. Specific Instance in a General Category			
Guggenheim Museum, NYC	Not found	Prince Charles	Not found
4. General			
Apple		Elephant	
5. Abstract or Affective			
A picture that represents success		A picture that represents tranquility	
6. General with Abstract or Affective Qualities			
A landscape that symbolizes freedom		A woman's face expressing joy	Not found
7. Iconological or Symbolic			
Patriotism		Santa Claus	
8. Object Pole			
Image to put on the cover of a camp brochure		Image to put on the front of a child's lunchbox	

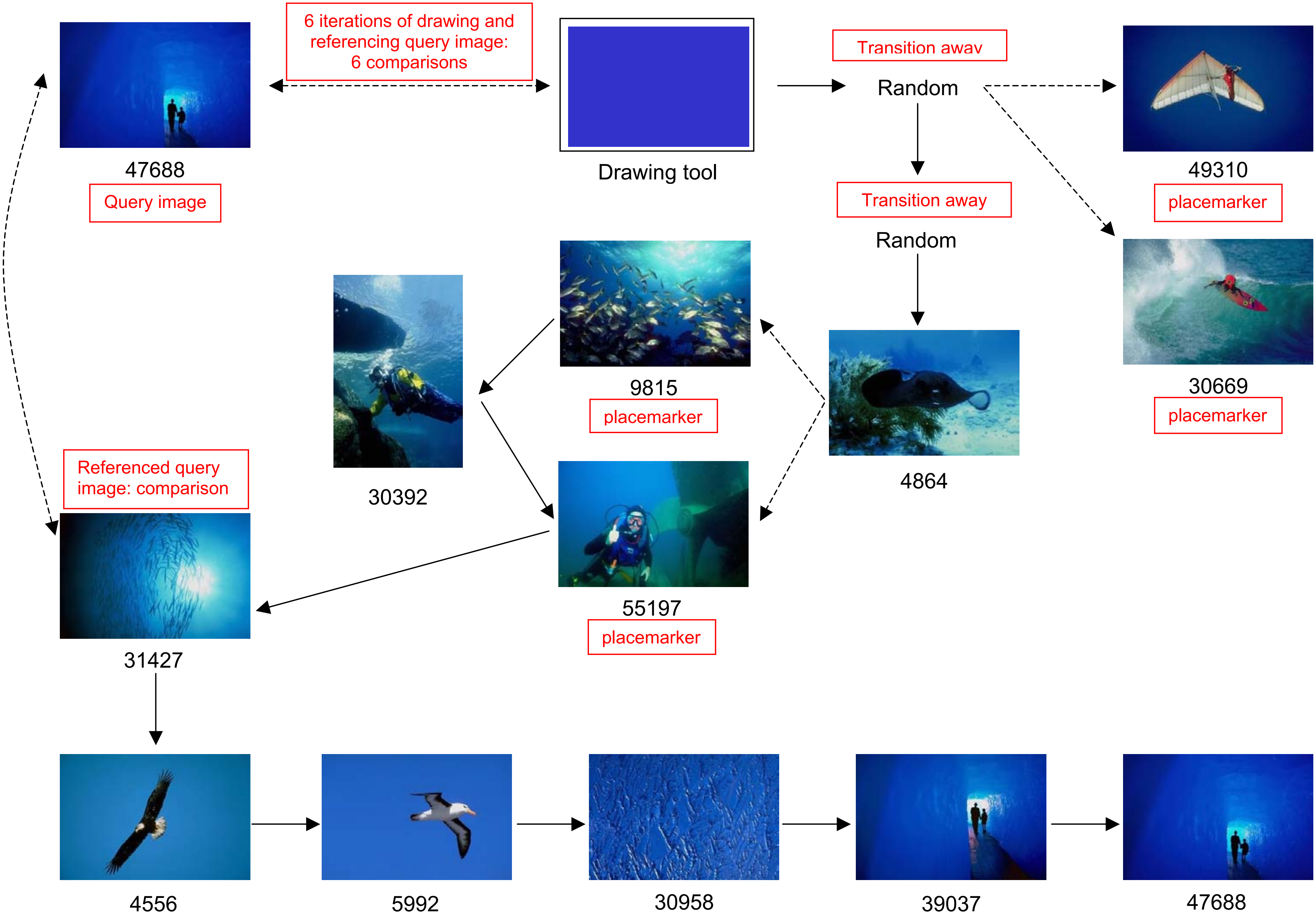
Appendix A: Queries, continued

Holley's Query	Image that satisfied query	Alison's Query	Image that satisfied query
9. Information Pole			
A map containing a body of water	Not found	A DNA double helix	Not found
10. Visual Content			
You are designing fabric, and you are looking for spherical objects with a prickly texture as a pattern inspiration		A generic scenic image for a university's publication that heavily features their school colors - yellow and green, and is of vertical orientation.	

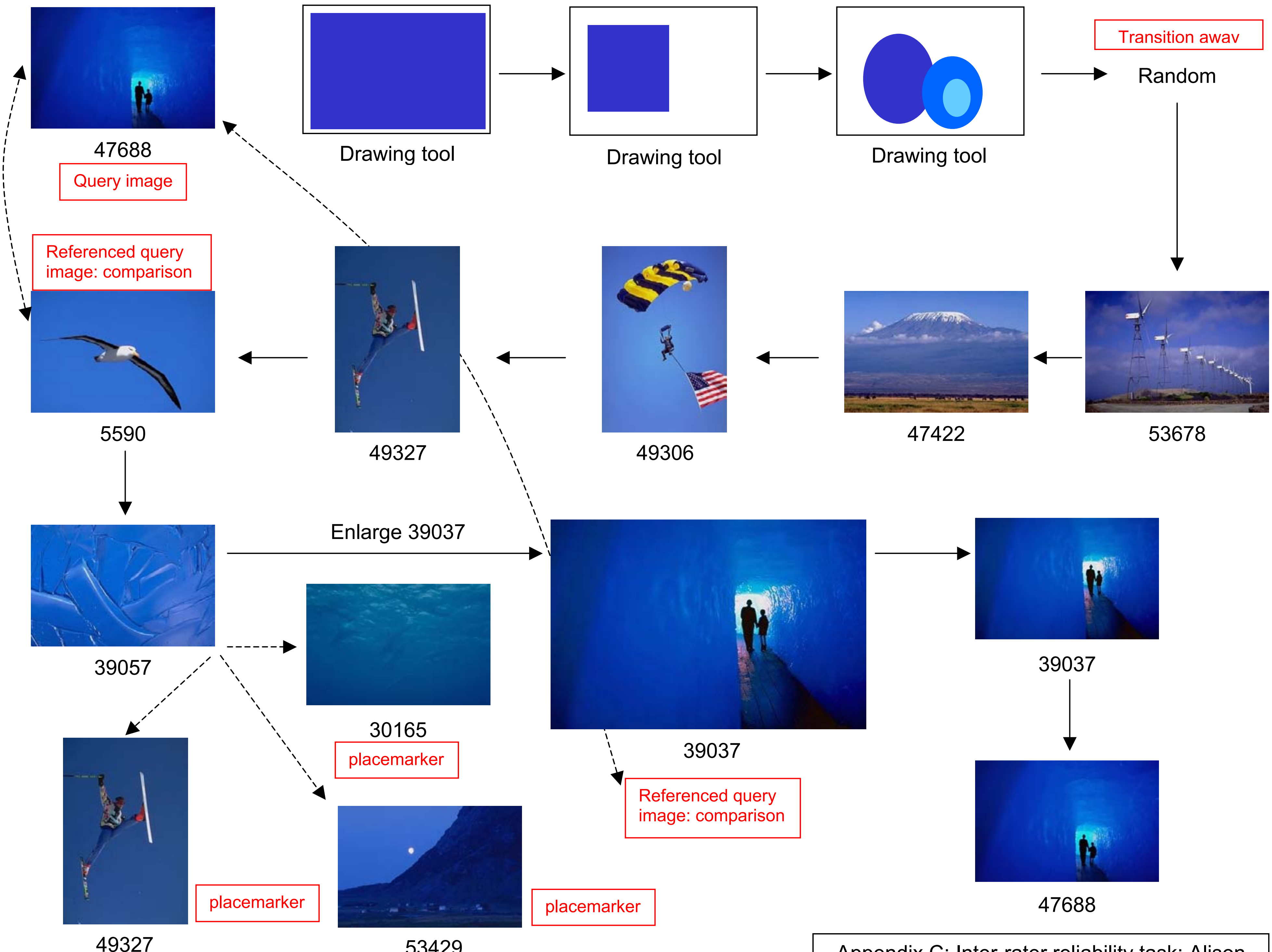
Task Number	4
Subject	Holley
Query	find an image of an apple
Task Entry Point	drawing tool
Date	4/20/2001
Start Time	10:14
End Time	10:18
Total Number of clicks	13
Initial Mental Image	thinking of a red apple against a very plain background - apple is the only object in the frame

		Identification and Place Marking			Orientation	Transitions		Anomalies	Comparisons
Click #	Task/User Action	Image # / place mark	Return?	Method	lost/dead end?	toward/away	comments	comments	comments
1	drawing tool								
2	submit	20430							
3	20430	45494				toward			it's an apple - but it's not exactly what I'm looking for
4	45494					toward			
5	random					away	try random just to see, knowing that I can come back to 20430		
6	random					away			
7	random					away			
8	random				dead end	away			
9	type 20430	45516	√	directly enter number		toward	same set as previous		
10	45516	38389							
11	enlarge 45516							I can't quite tell what this is from the small thumbnail.	
12	click back button								
13	38389								This isn't what I initially had in mind, but I like it. Once I began to see other pictures of fruit in a context, I started thinking in that mode instead of a single apple

Appendix B: Sample Data Collection Instrument

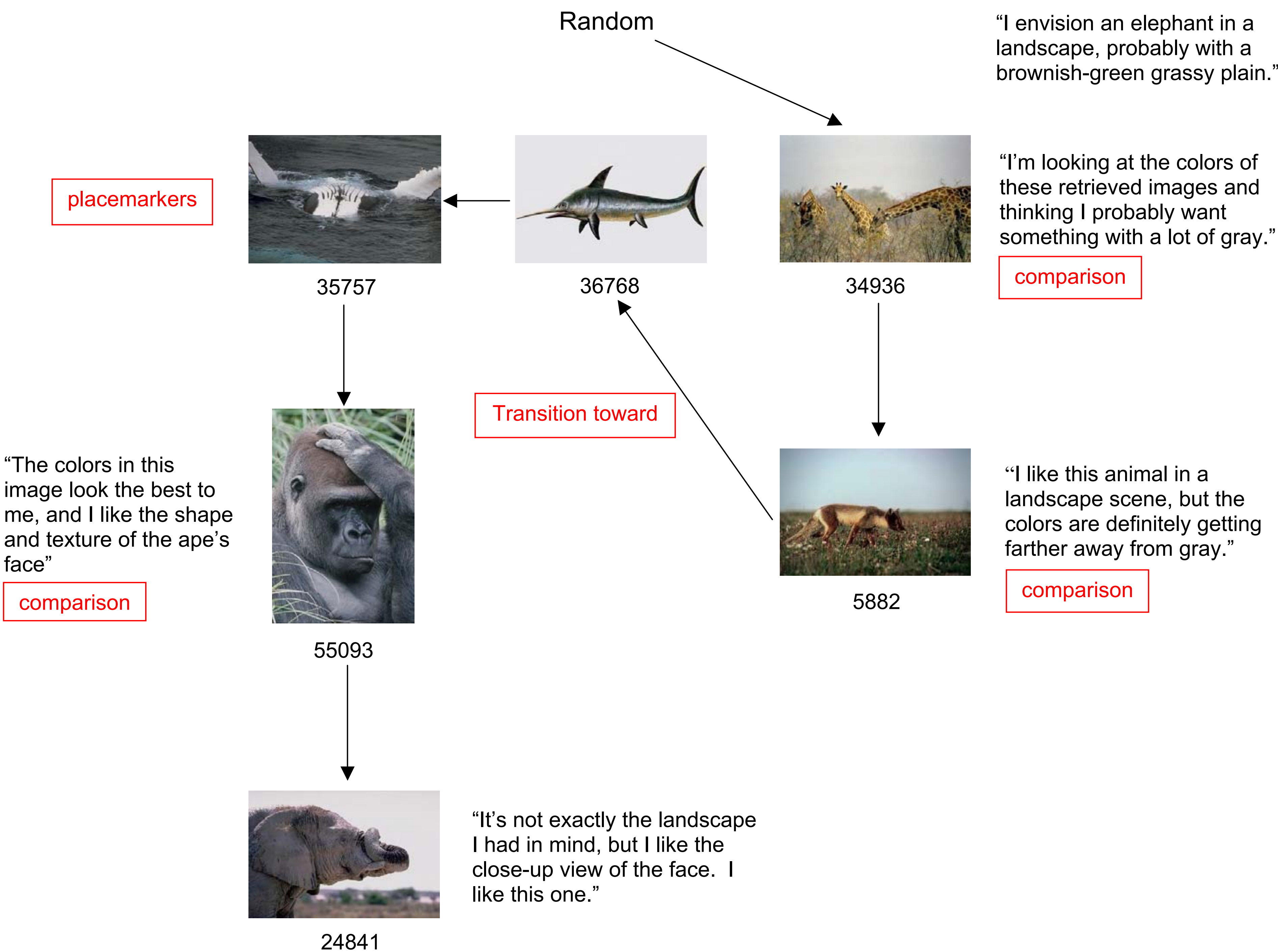


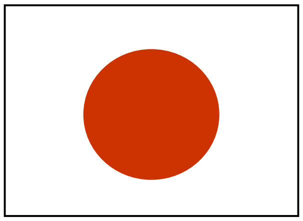
Appendix C: Inter-rater reliability task: Holley



Appendix C: Inter-rater reliability task: Alison

Appendix D: Anatomy of a Task (A4)





Drawing tool

"I am thinking of a red apple against a plain background"

20430

"It's an apple, but not exactly what I'm looking for."

comparison

placemarker

45494

Random

Random

Random

Random

Type "20430"

"I'm going to try random knowing that I can come back to 20430."

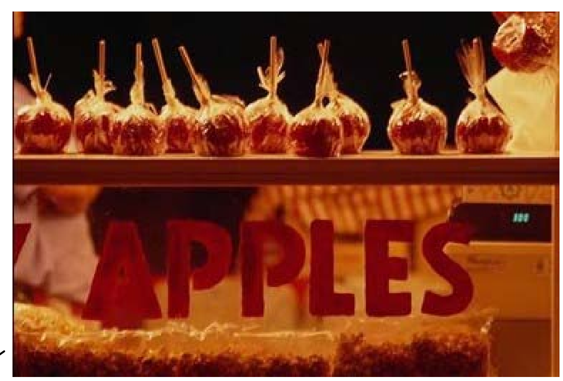
Transition away

Transition away

44551

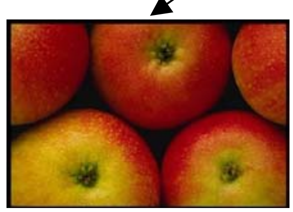
Enlarge 44551

"I can't quite tell what this is from the small thumbnail."



44551

Resolution of anomaly

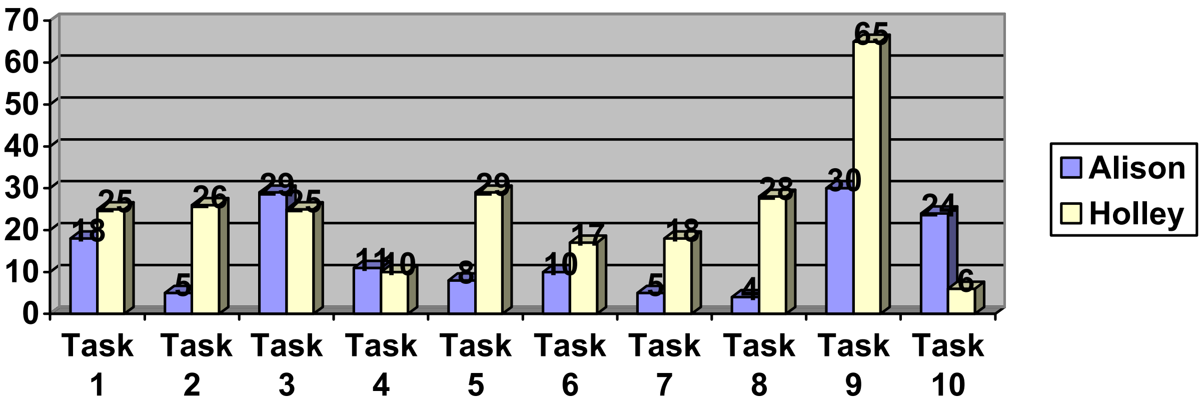


38389

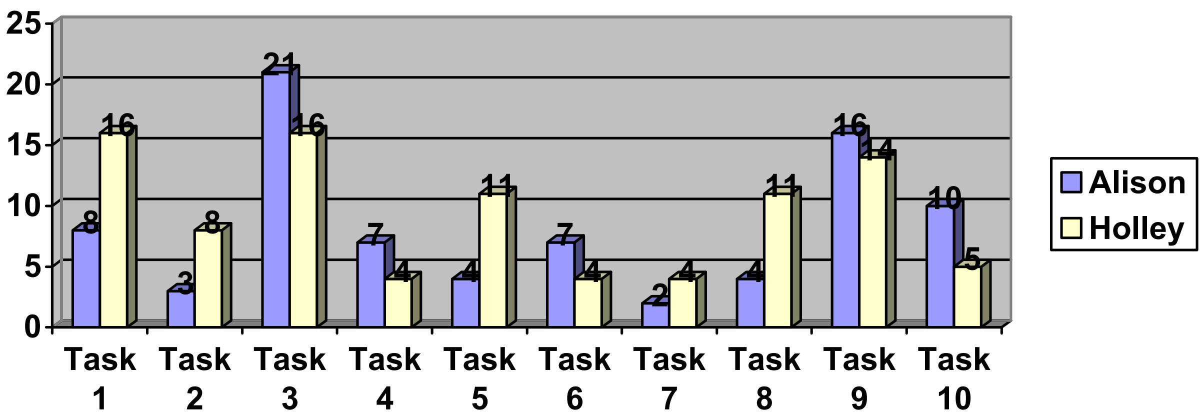
"This isn't what I initially had in mind, but I like it. Once I began to see other pictures of fruit in a context, I started thinking in that mode instead of a single apple."

Appendix E: Additional Data Analysis

NUMBER OF CLICKS PER TASK



PLACE MARKERS



TIME PER TASK IN MINUTES