01 Jun 2001

Image Retrieval Knowledge and Art History Curriculum in the Digital Age

Hsin-Liang Chen
Missouri University of Science and Technology, chenhs@mst.edu

Image Retrieval Knowledge and Art History Curriculum in the Digital Age

1 author:

Hsin-liang Chen
University of Massachusetts Boston
76 PUBLICATIONS **425** CITATIONS

SEE PROFILE

Image Retrieval Knowledge and Art History Curriculum in the Digital Age

Hsin-liang Chen

Graduate School of Library and Information Science, The University of Texas-Austin

chen@gslis.utexas.edu

ABSTRACT

This study investigates end-users' image queries by comparing the features of the queries to those identified in previous studies proposed by Enser and McGregor (1992), Jorgensen (1995) and Fidel (1997) in order to discover the utility of these existing features for the art history field and identify any expansions or new features. It also examines relationships between user search tasks and image query modes. Enser and McGregor's categories of Unique and Non-unique, and Jorgensen's classes of Location, Literal Object, Art Historical Information, People and People-related Attributes received high degrees of matching by three reviewers. This finding can be applied to add more details to Enser and McGregor's four categories (Unique, Non-unique, Unique with refiners, and Non-unique with refiners) and to re-group Jorgensen's 12 classes of image attributes.

This study also found several significant relationships between the participants' retrieval tasks and query modes. The participants who used fewer keywords from their topic title and topic description had a larger number of keywords or phrases they planned to use. A significant difference was found between the mean of the search keywords or phrases participants planned to use and the mean of the search keywords or phrases they actually used. The participants who submitted a higher number of keywords or phrases they planned to use were able to draw more pictures representing their topic title and the participants who had a greater number of keywords or phrases actually used also generated a larger number of post-search drawings. A significant relationship was also found between the level of success and the number of keywords or phrases participants planned to use. Implications for

curriculum design in the art history field, library instruction, image indexing tools and image retrieval system design are proposed.

INTRODUCTION

Images are used intensively in the field of art history. In a typical art history class, an instructor presents slides to students and they discuss the historical features of the slides. Instructors get slides from a departmental slide library and students go to the library to find related images from reference materials, journals or monographs. But computing technology is changing this scenario. Many museums and galleries digitize their collections and mount their digital images online. With these digital resources, researchers, educators and students in the art history field can retrieve images from collections around the world. However, these image resources apply different image management schema to control their digital collections and users have to acquire new knowledge to use these image resources. Before retrieving images, a user needs to know how to find appropriate images from different resources by using different skills and strategies. The new knowledge and skills of using digital image management systems have not been studied comprehensively.

Two major indexing approaches--concept-based and content-based--exist in the image retrieval field (Rasmussen, 1997; Chen & Rasmussen, 1999). Many previous studies on art image retrieval focused exclusively on the concept-based approach (Besser, 1990, 1991; Enser, 1995; Enser & McGregor, 1992; Hastings, 1995a, 1995b; Layne, 1994; Markey, 1984, 1988; Petersen, 1990; Tibbo, 1994; Walker & Thomas, 1990). Investigators from the computing technology community are interested in using a variety of approaches with algorithms and artificial intelligence techniques to index and retrieve images (Cawkell, 1994; Chang, 1989; Chang, et al., 1997; Faloutsos et al., 1994; Ma & Manjanath, 1999).

New technologies have placed the emphasis on automatic image indexing and content-based retrieval, although it is not clear how the retrieval functionality found in these

systems correlates with image information needs of real users. The use of image retrieval systems varies in different fields since users have their own specific information-seeking behavior, and need unique features designed for their tasks. A user-centered study can provide in-depth understanding of the user's information needs and his/her cognitive abilities. The understanding can be applied to design better user training, tutorials and system functions. In order to discover such understanding this project studied the user's image retrieval behavior and addressed implications for the curriculum design of the art history field.

BACKGROUND TO THE STUDY

Brilliant (1988) points out that most art historians can represent whole or partial art objects in their visual memory when searching art objects. Enser and McGregor (1992) also find that some patrons include a rough sketch to describe pictorially their requirements. Holt and Hartwick (1994a, 1994b) state that users search art images based on what things look like rather than using words to create descriptions using the QBIC (Query By Image Content) image retrieval system. However, Holt and Hartwick also find that users have problems in searching images by color and shapes because of the complexity of art images. Hastings (1995a, 1995b) investigates art historians' queries in using an art image database. She classifies the major features of the queries: identification, subject, text, style, artist, category, comparison, and color. Garber and Grunes (1992) point out that an art image retrieval system should allow the user to describe an image by entering explicit descriptions of the image. Alternatively, the user may target an identified image and use that as a source for finding similar images. Several previous studies provide basic understandings of image queries and retrieval tasks (Enser and McGregor, 1992; Enser, 1995; Jorgensen, 1995; Fidel 1997).

Enser and McGregor's study at the Hulton Deutsch Collection Ltd.

Enser and McGregor (1992) analyzed 2,722 image requests at the Hulton Deutsch Collection Ltd. The Hulton Deutsch Collection Ltd. houses over fifty distinct collections of images, the great majority of which are held in the form of negatives and prints. With approximately half a million engravings, woodcuts, drawings, cartoons and maps, the combined size of the collections is about ten million images.

Enser and McGregor classify image requests into four categories: 1) Unique, 2) Unique with refiners, 3) Non-unique, 4) Non-unique with refiners (Table 1). The property of uniqueness is readily applied to a request for the visual representation of an entity, the desired, particular occurrence of which can be differentiated from every other occurrence of the same entity type (Enser & McGregor, 1992, p.17). They point out that almost 70% of the requests were for a unique person, object or event, and 34% of the requests were refined (mostly by time).

Table 1. Categories of Enser & McGregor's 1992 study

Category	Example
Unique	George III
Unique with refiners	St. Paul's, 1920s-1950s
Non-unique	5-6 year old boy in silhouette
Non-unique with refiners	shell shock after First World War

Jorgensen's three image retrieval tasks and twelve image attribute classes

Jorgensen (1995) randomly selects seventy-seven images from *The Society of Illustrators 25th Annual of American Illustration,* and asks participants to do three tasks: 1) a describing task, 2) a searching task, and 3) a sorting task. In the describing task, forty-eight participants view 6 projected images and write a description of each. In the searching task, eighteen participants are given two terms representing abstract concepts, such as happy or mysterious, and then browse the set of seventy-seven images to find those relevant to the queries. In the sorting task, the same eighteen participants sort seventy-seven images into

groups for their own use as if the images are their personal collection. She classifies the verbal protocols into twelve classes of image attributes (Table 2).

In 1998, Jorgensen conducted another study with 107 first semester MLIS students with a variety of previous undergraduate majors (Jorgensen, 1998). Some of the 107 participants already had a graduate degree. The 107 participants viewed 6 projected color images which were randomly selected from *The Society of Illustrators 25th Annual of American Illustration*. The participants described the six projected images and Jorgensen classified the descriptions into the 12 classes of image attributes.

Table 2. Jorgensen's 12 classes of image attributes

Attribute Class	**Description**
• Literal object	• Named objects which are visually perceived, e.g. body parts, clothing
• People	• The presence of a human form
• People-related attributes	• The nature of the relationship among people, social status or emotions
• Art historical information	• Information related to the production context of the image, e.g. artists, medium, style
• Color	• Specific named colors or terms relating to various aspects of color
• Visual elements	• Elements such as composition, focal point, motion, shape, texture
• Location	• Both general and specific locations within the image
• Description	• Descriptive adjectives, e.g. wooden, elderly, or size or quantity
• Abstract concepts	• Attributes such as atmosphere, theme, or symbolic aspects
• Content/story	• A specific instance being depicted
• External relationships	• Relationships to attributes within or without the image, e.g. similarity
• Viewer response	• Personal reaction to the image

Fidel's Data and Objects Poles

Fidel (1997) proposes the Data Pole and the Objects Pole (Table 3). She points out that images are used as sources of information at the Data Pole. For example, a physician may need to use a slide of a normal foot to help decide if a patient's foot is flat. On the other hand, images are needed as objects at the Objects Pole. In this Pole, a slide librarian may be asked to search for slides that represent a specific idea or object. However, there are some in-between cases such as graphic artists and art historians who want to retrieve images both as information sources and as objects.

Table 3. Summary of Data Pole and Objects Pole

Data Pole	**Objects 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 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
• *Ofness* often equals *aboutness*	• O*fness* 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 the whole answer set

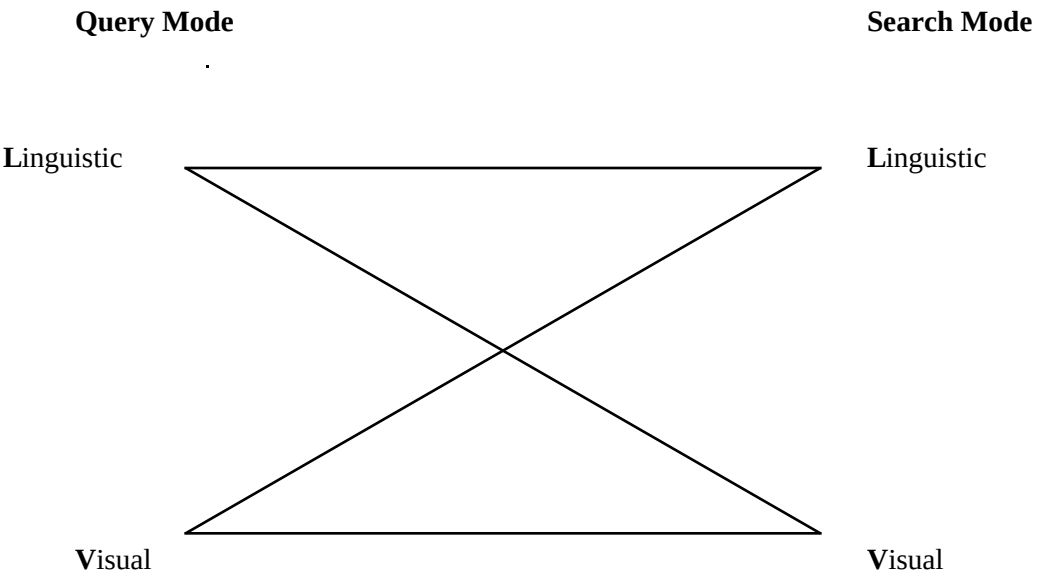
- Browsing is time consuming

- Browsing can be done rapidly

These three studies provide valuable classifications of users' image queries. However, some issues need to be addressed. For instance, users' feedback is not considered in the studies of Enser and McGregor, and Fidel. The queries are analyzed and judged by librarians or researchers. Do the images provided by these agencies satisfy the users? The answer is not available in their studies.

The studies of Enser (1995) and Jorgensen (1995) provide a valuable foundation in the analysis of art image retrieval tasks. Enser (1995) presents four models of image retrieval (Figure 1). They are: 1)linguistic query mode-linguistic search mode (LL); 2)visual query mode-visual search mode (VV); 3)visual query mode-linguistic search mode (VL); 4)linguistic query mode-visual search mode (LV).

Figure 1. Enser's Four Models of Image Retrieval



In the LL model, a query is expressed linguistically and a match is sought with the linguistic surrogates of the images in the collection. Surrogates can be the titles, authors, creators,

captions and index terms. Rasmussen (1997) calls this model the concept-based indexing approach. In the VV model, users are assumed to be able to submit a query in the "I want something which looks like this" form. Rasmussen calls this model the content-based indexing approach. Currently there are several systems available. These systems, IBM's QBIC (Query By Image Content), Visual Information Retrieval (VIR) Image Engine, and VisualSEEK, provide methods for retrieving digital still-images by using visual features of imagery, such as color, texture, shape and sketch, and spatial similarity. These functions are also used in combination with text and other related information (IBM, 1998; Virage, 1998; Advent, 1998). Enser points out that the VL model serves as an indexing paradigm rather than a retrieval strategy and uses the NASA Visual Thesaurus as an example. In one application involving astronomical images structures are specified as sets of pixels with specific values for their identifying attributes, such as color, morphological, and geometrical relations. Once these structures have been identified and named, their occurrences within newly-presented images can be sought by pattern-matching techniques, and their discovery within a particular image results in that image being indexed by the name of that structure (Enser, 1995, p. 158). In the LV model, the notion of digitized reference images or "hyper-icons" of the selected entities encoded within a chosen subject domain of a machine-readable knowledge classification scheme, such as Iconclass or Telclass. Each such image would be artificially modulated in order to generate the required number of images.

THE RESEARCH PROBLEM

The investigator was interested in the end-users' image queries and search tasks during the search process. The end users' queries were analyzed with Enser and McGregor's 4 categories; Jorgensen's 12 image attribute classes and Fidel's 2 poles.

Jorgensen classified the image retrieval tasks into three stages: describing, searching and sorting. However, Marchionini (1995) points out that a task includes an articulation,

usually stated as a question, and the mental and physical behaviors of interacting with search systems and reflecting on outcome. According to Marchionini, the sorting task should be considered as an interacting task. In this study, three retrieval tasks--describing; searching and interacting--were examined.

The objectives of this study are to: 1) map the features of the queries collected from end users with real information needs in the art history field; and 2) discover associations between user search tasks and query modes. The following questions were addressed:

A. Validation of the previous studies:

1. To what degree do the queries map to the previously identified image features: categories; image attribute classes; and poles?
2. Do these current features (categories, image attribute classes, and poles) need be modified or expanded in order to map image queries in the art history field?
3. Are totally new features required to describe queries in the art history field?

B. Task behavior and query modes:

1. Can any associations be identified between the three tasks (describing, searching and interacting) and query modes?
2. Is there a relationship between user characteristics (gender and information retrieval experience), the user's judgment on the degree of success of his/her search and the grade of the written paper?

RESEARCH DESIGN

Participants

The students in the two medieval art courses studied were required to finish a written paper to fulfill their course requirements. The instructor provided a list of topics within the themes of the classes. The instructor asked the students to include at least twenty images in

the paper, and their topics must be in the context of medieval art and approved by the instructor. Images could be either in digital or in hard-copy form, and image resources could be a library collection, a museum collection, museum publications such as postcards and catalogues, web sites on the Internet, or any other source. Students were encouraged to explore their searches in different environments.

Procedure

Two questionnaires, a pre-search questionnaire and a post-search questionnaire, were used in this study. The self-administered pre-search questionnaires were used to collect each user's image search topic, a description of the topic, and keywords or phrases which they planned to use in their searches. The participants were required to fill out the questionnaire and to return it to the investigator before they started their searches. The self-administered post-search questionnaire was used to collect keywords or phrases which participants used in their searches. The post-search questionnaire was collected with the written paper.

Post search interview

Prior to the post-search interview, the questionnaires were reviewed carefully by the investigator and each participant was asked to explain ambiguous answers or points. The participants were asked to fill out their demographic information and to clarify the answers given in the questionnaires. Chen (1999) described detailed research methodologies and data analysis methods in his dissertation.

Content analysis of verbal data

Three MLIS students with knowledge of indexing and/or classification were recruited from the Department of Library and Information Science at the University of Pittsburgh. The three reviewers were trained by the researcher and studied the findings of the three studies to

understand the meanings of these features. As part of the training, two participants' queries were randomly selected from the 29 participants. The three reviewers judged the two sets of queries and presented reasons for each judgement and the level of confidence about their judgement. By doing this, the researcher was able to verify that the reviewers understood the analysis procedure. The researcher also analyzed reasons given by the reviewers to generate a set of guidelines and explanations for each feature which facilitated the clarity of the whole mapping process.

Five hundred thirty-four queries were collected from the pre-search and post-search questionnaires. The reviewers agreed to treat each participant's topic title, title description and keywords as a whole set of data and made judgements based on the set of data. Table 4 presents an example of one participant's queries.

Table 4. Example of a participant's queries

Pre-search questionnaire
• **Topic Title** Venice, Italy • **Title description** Vince, Italy is a major city within Italy. It is a cultural center. Two of its major sites are Doges Palace and St. Mark's. It is also a major city of the Renaissance. • **Search keywords or phrases (plan to use)** 1. Vince, Italy 2. Renaissance Italy 3. Doges Palace 4. Italian sculpture 5. Italian culture 6. Italian cathedrals
Post-search questionnaire
• **Search keywords or phrases (actually used)** 1. Vince, Italy 2. Doges Palace 3. St Mark's Basilica 4. Gothic architecture 5. Byzantine architecture 6. Mosaics

RESULTS

Participants' characteristics

Eighteen students registered for the HAA 0240 class and twenty for the HAA 1200 class. Due to several personal factors, some students withdrew from the two courses. Eleven students in the HAA 0240 class and eighteen in the HAA 1200 class finished all the research requirements successfully: pre-search and post-search questionnaires, a post-search interview with the investigator, and the written paper.

Table 5 presents participants' characteristics in terms of academic status, gender, age and information retrieval (IR) experience. IR experience was judged by the frequency of use of online public access catalogs (OPACs) and electronic databases. Table 6 and 7 show the distribution of number of the keywords and drawings submitted by the two courses of participants before and after their searches. Three participants changed their topics after completing the pre-search questionnaires. These were excluded from the analysis of task behavior and query modes.

Table 5. Participants' characteristics (N=29)

Characteristics	Measurement	Number of participants	%
Academic Status	1. Sophomore	3	10.35
	2. Junior	5	17.24
	3. Senior	16	55.17
	4. Post-bachelor	2	6.90
	5. Masters	3	10.35
Gender	1. Male	13	44.83
	2. Female	16	55.17
Age	1. Under 20	2	6.90
	2. 20-25	20	68.97
	3. 26-30	3	10.35
	4. 31-35	1	3.45
	5. 36-40	2	6.90
	6. 41-45	0	0.00
	7. 46-50	0	0.00
	8. above 50	1	3.45
IR Experience			

OPACs	1. Daily	4	13.79
	2. Weekly	9	31.04
	3. Monthly	11	37.93
	4. Less than monthly	5	17.24
Electronic databases (other than OPACs)	1. None	14	48.28
	2. Daily	1	3.45
	3. Weekly	0	0.00
	4. Monthly	5	17.24
	5. Less than monthly	9	31.04

Table 6. Distribution of number of queries

Course	Pre-search				Post-search			
	Min	Max	Mean	SD	Min	Max	Mean	SD
HAA 0240 (N=11)	5	11	7.27	1.79	4	21	10.91	5.45
HAA 1200 (N=18)	3	11	7.17	2.35	4	19	8.17	3.82

Table 7. Distribution of number of drawings

Course	Pre-search				Post-search			
	Min	Max	Mean	SD	Min	Max	Mean	SD
HAA 0240 (N=11)	1	5	2.55	1.37	2	6	4.18	1.25
HAA 1200 (N=18)	0	6	2.72	1.67	2	5	3.00	0.91

Analysis of image queries

Enser and McGregor's four categories

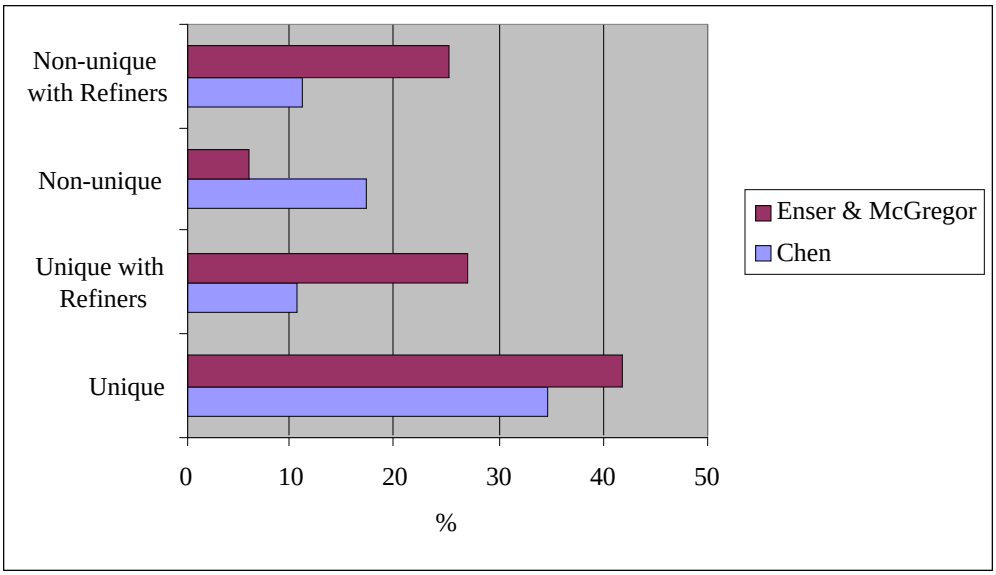
Seven hundred seventeen judgments were made for Enser and McGregor's four categories and 526 (73.37%) of these judgments were effective. The Unique category had 249 (34.73%) effective judgments and the Non-unique category received 113 (17.16%) effective judgments. On the other hand, the Unique with Refiners category received 75 (10.46%) effective

judgments and the Non-unique with Refiners category had 79 (11.02%) effective judgments. The degrees of matching of the four categories vary from 6.82 to 5.84 on a 7-point scale.

However, the three reviewers had some conflicts between the Unique and Non-unique categories. Sometimes they also interpreted a refiner differently. The reasons for the different judgments are hidden meanings of the queries and personal subjective judgment.

This mapping procedure reveals that these four categories were agreed upon by either two or three of the reviewers with high mean degrees of matching, especially for the Unique and Non-unique categories. However, the significance of the four categories differs from the results found in Enser and McGregor's 1992 study (Figure 2).

Figure 2. Distribution of categories in Enser and McGregor's study and in Chen's study



In this study, the use of an intermediary was not a factor controlled by the investigator. The results of the mapping process showed that the participants seldom added refiners such as a specific time period, a certain geographical term or a special material. One possible reason was that the participants were looking for relevant medieval images in the European countries which were covered in the two courses. However, it was not obvious why most queries did not include refiners and why the participants often used non-unique queries.

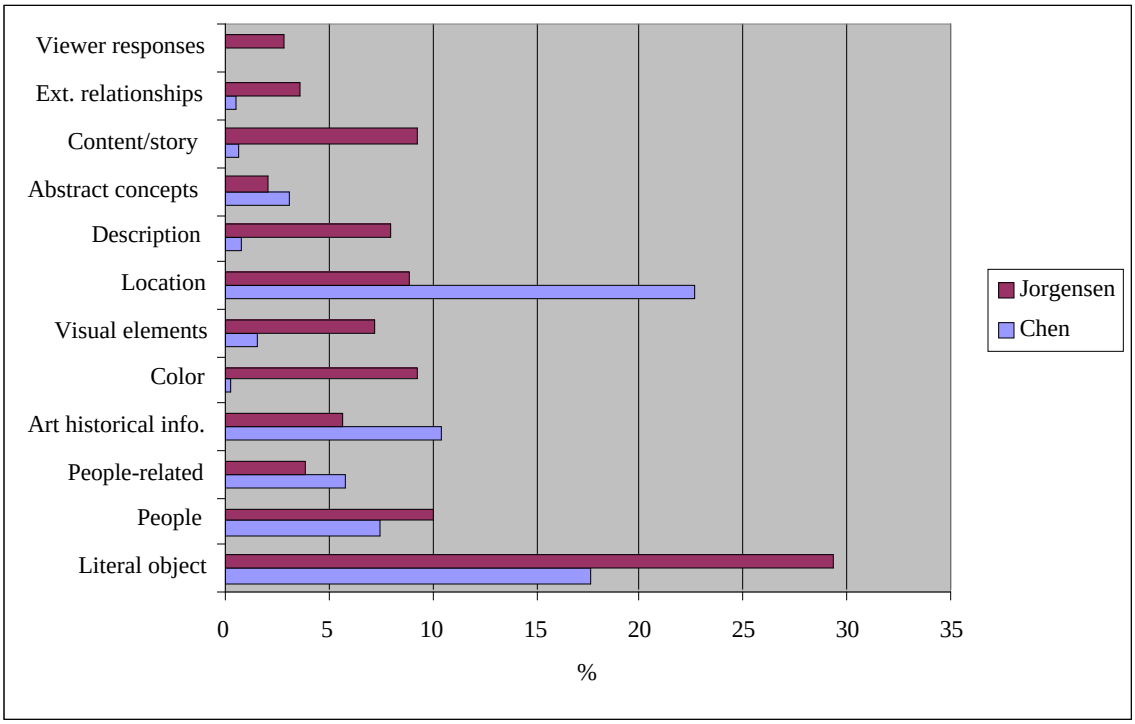
Jorgensen's 12 classes of image attributes

The total number of judgments made for Jorgensen's 12 classes of image attributes was 1207 and 857 (71%) of these judgments were effective. Only 5 classes received a substantial number of effective judgments which was over 5%. Based on percentage of effective judgments, they are: Location (22.62%), Literal Object (17.65%), Art Historical Information (10.52%), People (7.46%), and People-related Attributes (5.80%).

On the other hand, the Color, Visual Elements, Description, Abstract Concepts, Content/story, External Relationships, and Viewer Response classes did not receive a significant number of effective judgments. Lowest were the Color class which received only 4 effective judgments agreed upon by two reviewers, and the Viewer Response Class which did not have any effective judgments at all. Regarding the degree of matching, the 12 classes had a larger range (from 6.64 to 4.77), than those of Enser and McGregor's four categories (from 6.82 to5.84).

The findings of this study and Jorgensen's two studies indicate that some of the 12 classes of image attributes such as the People, Location and Literal Objects classes have a consistently high level of occurrence in participants' descriptions or queries. Basically, this investigator and Jorgensen both identified these features based on participants' queries and descriptions (Figure 3). In order to modify the classes which have a low level of occurrence or to discover new features of image queries, other methodologies for data collection should be considered. For example, a thinking out loud method may provide useful data on how a user changes his/her mind when describing an image information need. A transaction log analysis may find an evolving process of the construction of a query. These methodologies should be considered for further studies.

Figure 3. Distribution of categories in Jorgensen's study and in Chen's study



Fidel's Data Pole

The total number of judgments made for Fidel's Data Pole was 1207 and 908 (51.84%) of these judgments were effective. The first feature (Images provide information) received 299 (15.54%) effective judgments; the ninth feature (Biographical attributes are not likely to play a role) had 272 (14.14%) effective judgments; the tenth feature (To satisfy requests may require sets of more than one image) had 131 (6.86%) effective judgments; and the twelfth feature (Browsing is time consuming) received 225 (11.69%) effective judgments. On the other hand, the other 8 features of the Data Pole did not receive a significant number of effective judgments.

Fidel's Objects Pole

The total number of judgments made for Fidel's Objects Pole was 1197 and 633 (52.88%) of these judgments were effective. The first feature (Images are objects) received 211 (17.62%)

effective judgments; the sixth feature (No evidence exists that color, shape and texture are important for retrieval) had 266 (22.22%) effective judgments; and the ninth feature (Biographical attributes are important for relevance assessment) had 113 (9.44%) effective judgments. On the other hand, the other 9 features of the Objects Pole did not receive significant numbers of effective judgments.

Regarding the number of effective judgments for Fidel's Data Pole and Objects Pole, the investigator has suggested other research methods for data collection because some features are user-related. These user-related features could only be judged by participants themselves. Chen (2001a) reports the detailed research result and analysis.

An analysis of image retrieval tasks

The second objective of the study is to discover associations between the three tasks (describing, searching and interacting) and query modes, and relationships between user characteristics (gender and IR experience), the user's judgment on the degree of success of his/her search and the grade of the written paper. Three participants changed their topics after completing the pre-search questionnaires. These were excluded from the analysis of task behavior and query modes.

The following questions were addressed:

Does the user's gender have any influence on his/her retrieval tasks?

The findings indicated that female participants used OPACs more often than the male participants in this study. No other relationships were found between gender and the other study parameters.

Does the user's experience with OPACs and electronic databases affect his/her search behavior?

Experience with OPACs was related to the level of confidence for post-search

drawings and the level of success for the search results. On the other hand, no relationship was found between the experience with electronic databases and the other study parameters.

Does the user use mostly keywords or phrases from his/her topic description?

The findings of this study found only 33.66% of the pre-search keywords or phrases were drawn from the topic titles and descriptions by the 29 participants. There are two possible reasons for the findings: most participants did not know how to describe their information needs, and/or they did not know how to extract keywords from the topic and description.

Pearson correlation coefficients showed there was a significant relationship between the describing task and the number of search keywords or phrases participants planned to use ($r=-0.36$ & $p=0.07$). This indicated that the participants who had a lower percentage of keywords or phrases drawn from their topic title and topic description had a larger number of pre-search keywords or phrases.

Is the user able to represent his/her search topic both in the linguistic and visual modes?

The findings of this study found no significant relationships between the percentage of keywords or phrases drawn from the topic title and topic description and either the number of pre-search drawings or the level of confidence for the pre-search drawings. This study did not find any evidence which showed any relationship between the describing task in the linguistic mode and the describing task in the visual mode.

The findings of this study also found no significant relationships between the number of search keywords or phrases participants actually used in the search and the number of pre-search drawings, or between the number of search keywords or phrases participants actually used in the search and the mean score of the level of confidence for the drawings. This study did not find any evidence which showed any relationship between the interacting task in the linguistic mode and the interacting task in the visual mode.

When the user generates more textual queries, does he/she feel more confident about

drawing a picture to represent his/her search topic?

Pearson correlation coefficients found a significant relationship between the number of keywords or phrases participants planned to use and the number of pre-search drawings ($r=0.43$ & $p=0.03$). This indicated that the participants who generated more keywords or phrases which they planned to use were able to sketch more drawings representing their topic title.

Pearson correlation coefficients found a significant relationship between the number of the keywords or phrases actually used and the number of post-search drawings ($r=0.42$ & $p=0.03$). This indicated that the participants who generated more keywords or phrases during their searches were able to produce more drawings representing their topic title.

However, Pearson correlation coefficients found a significant relationship between the mean of pre-search drawings and the mean score of the level of confidence for these drawings while the same result could not be found between the mean of post-search drawings and the mean score of the level of confidence for these drawings. It indicated that the participants who produced more drawings after the interacting task were not necessarily more confident than participants who produced fewer drawings.

Regarding the level of confidence, the 26 participants gave relatively low scores both on the searching (2.90) and interacting (3.80) tasks based on a 7-point scale (1, extremely unsuccessful, to 7, extremely successful). This finding showed that the majority of the 26 participants did not feel at all confident about their drawings representing their topic title either before or after their searches.

The 26 participants reported their reasons for the scores of the degree of matching: the majority of the participants (21; 80.77%) did not consider themselves as having good drawing skills either before or after their searches; some images were considered too complex to draw, and some participants were not sure about the details of the topic and/or the details of the images before the search.

After an interacting task, is the user able to represent his/her search topic both in the linguistic and visual modes?

Results of a t-test (t=2.96, p=0.01) and one-way ANOVA (F=8.75, p=0.01) found that the number of search keywords or phrases (9.42) participants actually used on the interacting task was significantly greater than the number of search keywords or phrases (7.00) participants planned to use at the 0.01 level. There was also a significant relationship between the number of search keywords or phrases participants planned to use and the number they actually used (r=0.465, p=0.02). This indicates that participants who submitted more keywords or phrases before their searches tended to use more keywords or phrases on the interacting task.

Results of a t-test (t=2.27, p=0.03) and one-way ANOVA (F=5.17, p=0.03) showed that, on the average, participants produced more drawings after their searches (mean=3.46) than before their searches (mean=2.65). However, there was no significant relationship between the number of drawings before and after, indicating that individual students who generated a relatively large number of drawings before their searches did not necessarily produce a relatively large number after their searches.

In addition, results of a t-test (t=2.20, p=0.04) and one-way ANOVA (F=4.82, p=0.04) also found that, on the average, students had a higher degree of confidence for the drawings after their searches (mean=3.80) than before their searches (mean=2.90). However, individual students who had a relatively higher degree of confidence for drawings before their searches did not necessarily have a relatively higher degree of confidence for drawings after their searches.

Does the user finish his/her search successfully?

Overall, the participants who gave a higher score on the level of success received a better grade of the written paper from the instructor. This showed an agreement between the participants' self-evaluation on their search results and the grade given by the instructor.

The findings of this study found that the participants' levels of success generally agreed with the grade on their written paper given by the instructor. Pearson correlation coefficients found a significant relationship ($r=0.33$ & $p=0.10$) between the mean score of the level of success (5.73, based on a 7-point scale) and the mean of the grade on the written paper (3.36, based on a 4-point scale). However, there was no significant relationship between the grade on the written paper and the other study parameters.

The 26 participants reported their reasons for the score of the level of success for search results:

- The majority of the participants (25; 96.15%) retrieved twenty or more images for the written paper. (The one participant who did not retrieve enough images reported that the book she needed was missing from the library.)
- The same 25 participants reported that the retrieved images were related to their topics, and
- All of the participants reported the search process was not difficult.

The investigator also interviewed the instructor for the grading criteria of the written paper. The instructor described her grading criteria as follows:

1. Breadth of images: whether the twenty images covered the range of the topic proposed by the student.
2. References: whether the student provided adequate citations to appropriate resources.
3. Conceptual structure: whether the framework of the written paper was presented in a strong logical order.
4. Writing: whether the student organized his/her ideas using text, images and references.

The instructor graded the participants' papers after the investigator interviewed the participants and she returned the papers to the participants before the semester ended. Besides the grading criteria, the instructor also pointed out that the students reported to her that they had problems drawing pictures to represent their topics due to the complexity and

ambiguity of their topics. Chen (2001b) reports the detailed research result and analysis.

DISCUSSION AND CONCLUSION

In the two courses, HAA 0240 and HAA 1200, slides were frequently used by the instructor. She asked students to point out any differences between paired slides. In general, students identified the contextual meanings of symbols, objects, colors, locations, materials, and contents within each slide. They also identified any associations of these above elements between slides. In spite of the highly visual mode of instruction, the participants did not have a strong level of confidence that their drawings represented their topics either before or after their searches. This finding showed that the majority of the 26 participants did not feel confident at all about their drawings representing their topic either before or after their searches.

This indicates that drawing should be considered in the design of art history curriculum. Students should be able to present art objects both in a textual statement and in a visual form which can be constructed by simple geometric shapes or complex sketches, and these skills should be considered as components of visual literacy.

With better drawing skills, students would be able to represent art objects for which they are looking. Further, they could generate their queries in a visual mode which is an extremely important search function for image retrieval currently and in the future. IBM's QBIC is a good example. This system allows users to draw different geometric shapes such as a circle, square or triangle to present an image query. The user can also generate a more complex sketch to present the search topic. The role of instruction is a vital component in the research of image retrieval. Improvements should not only rely on the invention of new image retrieval systems and indexing tools. Instruction, training and other human effort should be taken into account.

Regarding image retrieval knowledge, the significant relationship between

experience with OPACs and the level of success for search results indicates the importance of personal information infrastructure in the satisfaction of a user's information needs. In this study, IR experience was defined by two components: experience with OPACs and experience with electronic databases. Both of the two components are related to library instruction (Aluri, Kemp, & Boll, 1991). There was no significant difference between experience with electronic databases and the other study parameters. This showed the experience (or lack thereof) with electronic databases did not seem to play an important role in this study.

For future studies, the investigator is interested in the effect of the search functions/devices of the image retrieval systems on the user search behavior. The current search functions need to be studied with human cognitive models to discover: whether people can describe their image information needs based on the current search functions such as color, shape and location; whether the current functions are enough for the expression of image queries; and whether new search functions/input devices are needed and what they should be.

REFERENCES

Aluri, R., Kemp, D. A., & Boll, J. J. (1991). Subject analysis in online catalogs. Englewood, CO: Libraries Unlimited.

Advent. (1998). [On-line] Available: http://www.ctr.columbia.edu/advent/home-full.html

Besser, H. (1990). Visual access to visual images: The UC Berkeley image experience with electronic databases project. Library Trends, 38(4), 787-789.

Besser, H. (1991). Imaging: Fine arts. Journal of the American Society for Information Science, 42(8), 589-596.

Biederman, I. (1987). Recognition by components: A theory of human image understanding. Psychological Review, 94, 115-147.

Brilliant, R. (1988). How an art historian connects art objects and information. Library Trends, 37(2), 120-129.

Cawkell, A. E. (1994). A guide to image processing and picture management. Aldershot, Hampshire (UK): Gower Publishing Limited.

Chang, S. K. (1989). Principles of pictorial information systems design. Englewood Cliffs, NJ: Prentice-Hall.

Chang, S., Smith, J. R., Beigi, M., & Benitez, A. (1997). Visual information retrieval from large distributed online repositories. Communications of the ACM, (40)12, 63-71.

Chen, H. (1999). An analysis of image queries and image retrieval tasks in the field of art history. Unpublished doctoral dissertation. University of Pittsburgh.

Chen, H. (2001a). An analysis of image queries in the field of art history. Journal of the American Society for Information Science and Technology, 52(3), 260-273.

Chen, H. (2001b). An analysis of image retrieval tasks in the field of art history. Information Processing and Management, (37)5, 701-720.

Chen, H. & Rasmussen, E. M. (1999). Intellectual access to images. Library Trends, (48)2, 283-288.

Enser, P.G.B. (1995). Progress in documentation pictorial information retrieval. Journal of Documentation, 51(2), 126-170.

Enser, P.G.B., & McGregor, C.G. (1992). Analysis of visual information retrieval queries. Report on Project G16412 to the British Library Research and Development Department. London: British Library.

Faloutsos, C., Barber, R., Flickner, M., Hafner, J., Niblack, W., Petkovic, D., & Equitz, M. (1994). Efficient and effective querying by image content. Journal of Intelligent Information Systems, 3, 231-262.

Fidel, R. (1997). The image retrieval task: Implications for the design and evaluation of image databases. The New Review of Hypermedia and Multimedia, 3, 181-199.

Garber, S. R., & Grunes, M. B. (1992). The art of search: A study of art directors. CHI Human Factors in Computing Systems, 157-163.

Hastings, S. K. (1995a). An exploratory study of intellectual access to digitized art images. Proceedings of the 16th National Online Meeting, 177-185.

Hastings, S. K. (1995b). Query categories in a study of intellectual access to digitized art images. Proceedings of the 58th Annual Meeting of the American Society for Information Science, October 9-12, 1995, Chicago, IL. Medford, NJ: ASIS. 3-8.

Holt, B., & Hartwick, L. (1994a). Retrieving art images by image content: The UC Davis QBIC project, ASLIB-Proceedings, 46(10), 243-248.

Holt, B., & Hartwick, L. (1994b). `Quick, Who Painted Fish?'; Searching a Picture Experience with electronic databases with the QBIC Project at UC Davis. Information Services & Use, 14, 79-90.

IBM. (1998). [On-line] Available: http://wwwqbic.almaden.ibm.com/

Jorgensen, C. (1995). Image attributes: An investigation (Indexing systems, retrieval systems, computerized). Unpublished doctoral dissertation. Syracuse University.

Jorgensen, C. (1998). Attributes of images in describing tasks. Information Processing & Management, 34(2/3), 161-174.

Layne, S. S. (1994). Some issues in the indexing of images. Journal of the American Society for Information Science, 45(8), 583-588.

Ma, W. & Manjanath, B. (1998). Netra: A toolbox for navigating large image databases. Proceedings of IEEE International Conference on Image Processing (ICIP97), 1, 568-571.

Marchionini, G. (1995). Information seeking in electronic environments. Cambridge: Cambridge University Press.

Markey, K. (1984). Interindexer consistency tests: A literature review and report of a test of consistency in indexing visual materials. Library & Information Science Research,

6(2), 155-177.

Markey, K. (1988). Access to iconographical research collections. Library Trends, 37(2), 154-174.

Petersen, T. (1990). Developing a new thesaurus for art and architecture. Library Trends, 38(4), 644-658.

Rasmussen, E. M. (1997). Indexing images. Annual Review of Information Science and Technology, 32, 169-196.

Tibbo, H. R. (1994). Indexing for the humanities. Journal of the American Society for Information Science, 45(8), 607-619.

Virage, Inc. (1998). [On-line] Available: http://www.virage.com/market/vir.html/

Walker, F. L., & Thomas, G. R. (1990). Access techniques for document image experience with electronic Experience with electronic databases. Library Trends, 38(4), 751-786.