An analysis of image retrieval tasks in the field of art history

Hsin-liang Chen [*]

Graduate School of Library and Information Science, The University of Texas at Austin, Austin, TX 78712, USA

Abstract

This project applied Jorgensen's three image retrieval tasks (1995) and Enser's four models of image retrieval (1995) to investigate end-users' image retrieval tasks and image query modes in the field of art history. Twenty-six undergraduate students majoring in art history were recruited in this study. They were required to finish a term paper including at least 20 images. The participants' image retrieval tasks were collected by pre- and post-search questionnaires, and a post-search interview. There was a significant difference between the mean number of search keywords or phrases participants planned to use and the mean number of search keywords or phrases they actually used. The study found a significant relationship between the level of success for the search results and the percentage of search keywords or phrases drawn from the topic title or topic description participants planned to use. Implications for image indexing tools and image retrieval system design are proposed. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Image retrieval; Retrieval tasks; User studies; Art history

1. Introduction

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; Case, 1996; Enser, 1995; Enser & McGregor, 1992; Hastings, 1995a,b; Layne, 1994, 1997; Markey, 1984, 1988; Petersen, 1990; Tibbo, 1994; Walker & Thomas, 1990). Stam (1989) points out that there were two categories of studies on the art historians' information needs: studies of art historians as they sought or used information, and studies that analyzed the information that had been used in the art historians' published research. Most previous studies on art historians' information-

[*] Tel.: +1-512-232-9219; fax: +1-512-471-3971.

E-mail address: chen@gslis.utexas.edu (H. Chen).

seeking behavior and needs focused on the text-related information (Bates, Wilde, & Siegfried, 1993; Wiberley, 1988).

The use of image retrieval systems varies in different fields since users have specific information-seeking behavior, and need unique features designed for their tasks. Important issues such as user search behavior analysis and specific domain applications (Dervin & Nilan, 1986) need to be addressed to further enhance studies on image retrieval systems. New technologies have placed the emphasis on automatic image indexing and content-based retrieval, although it is not clear how the retrieval functionality of these systems correlates with image information needs of real users.

Enser (1995) proposes four models of pictorial information retrieval based on whether the query and search are presented in text (linguistic) mode or in image (visual) mode: (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). Jorgensen (1995) identifies three tasks: (1) a describing task, (2) a searching task and (3) a sorting task. In order to provide more suggestions for the design of art image retrieval systems, the investigator focused on user image retrieval tasks and queries in the field of art history. This study examined users' image retrieval tasks and queries, and addressed their implications for art image indexing tools and retrieval systems.

2. Background of the study

Brilliant (1988) points out that most art historians can represent whole or partial art objects in their visual memory when searching for art objects. Enser and McGregor (1992) also find that some patrons include a rough sketch to describe their requirements pictorially. Holt and Hartwick (1994a,b) state that users search for art images based on what things look like rather than using words to create descriptions using the query by image content (QBIC) image retrieval system. However, Holt and Hartwick also find that users have problems in searching for images by color and shapes because of the complexity of art images. Hastings (1995a,b) investigates into 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. The complexity of queries influences art historians as they select access points and computer manipulation techniques. 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. These findings suggest the need for further research into art image retrieval. The studies of Enser (1995) and Jorgensen (1995) provide a valuable foundation for the analysis of art image queries.

2.1. Enser's four models of pictorial information retrieval

Enser (1995) presents four models of image retrieval (Fig. 1). They are:
- linguistic query mode–linguistic search mode,
- visual query mode–visual search mode,
- visual query mode–linguistic search mode,
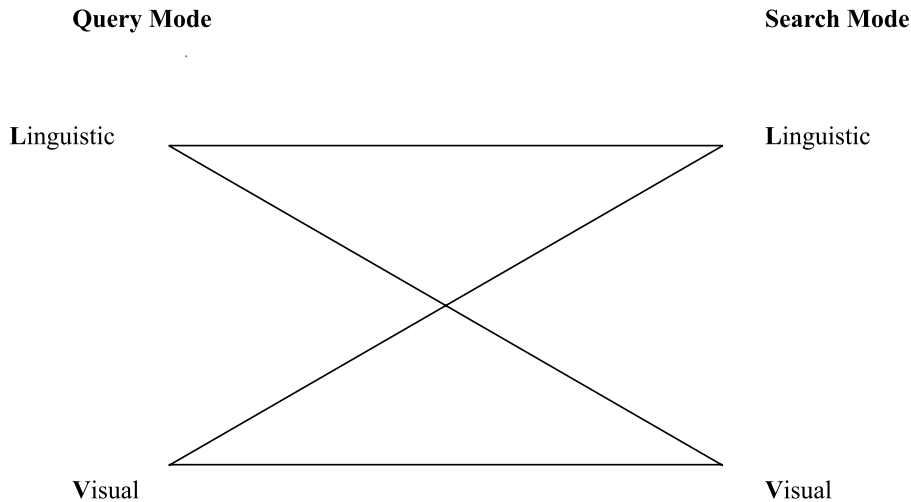- linguistic query mode–visual search mode.

Query Mode   **Search Mode**

Linguistic   Linguistic

Visual   Visual

Fig. 1. Enser's four models of image retrieval.

2.1.1. The model of LL

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 titles, authors, creators, captions and index terms. Rasmussen (1997) calls this model the concept-based indexing approach. LIS researchers have continued the tradition of constructing text-based indexing tools (Besser, 1990, 1991; Enser, 1995; Enser & McGregor, 1992; Hastings, 1995a,b; Layne, 1994; Markey, 1984, 1988, 1990; Tibbo, 1994; Walker & Thomas, 1990).

2.1.2. The model of VV

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. Several query-by-visual-example image retrieval systems have been reported in specialized subject areas (Beard, 1991; Cawkell, 1994; Chang, 1989; Chang, Smith, Beigi, & Benitez, 1997; Faloutsos et al., 1994; Gupta, Santini, & Jain, 1997). These systems like the IBM's QBIC, 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, Inc., 1998; Advent, 1998).

2.1.3. The model of VL

Enser explains 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 astronomical image 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).

2.1.4. The model of LV

In the LV model, digitized reference images or ''hyper-icons'' of the selected entities are 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.

Heidorn (1997) constructs a natural language processing system which can interpret text-based descriptions of physical objects and create three-dimensional graphic models of the objects. The system integrates current theories of human linguistic and visual processing, and embodies principles for the representation and processing of objects and relationships that permit communication between verbal and visual subsystems.

2.2. Jorgensen's three image retrieval tasks

Jorgensen (1995) randomly selects 77 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, 48 participants view six projected images and write a description of each. In the searching task, 18 participants are given two terms representing abstract concepts, such as happiness or mystery, and then browse the set of 77 images to find those relevant to the queries. In the sorting task, the same 18 participants sort the 77 images into groups for their own use as if the images were their personal collection.

These two studies provide a valuable insight into users' image retrieval tasks. However, some issues need to be addressed. Dervin (1989) points out that a balanced and complete approach to users is necessary if the system design is to be successful. Allen (1996) states that an information-as-process approach should be involved in the design of IR systems, and argues that user-centered design focuses on information-as-process, particularly on the ways that information systems meet the information needs of users. The studies reveal that the importance of user-centered issues does not get enough attention in information retrieval in general. To try and address this, this research project focuses on end-users' perspectives and investigates their image information needs. It is hoped that the results will reflect users' true needs in searching for images and will provide valuable evidence for the features identified in the previous studies.

3. The research problem

Jorgensen defines three tasks (describing, searching and sorting) in retrieving images. Marchionini (1995) points out that a task includes an articulation, usually stated as a question, the mental and physical behaviors of interacting with search systems, and reflecting on the outcome. According to Marchionini, the sorting task should be considered an interacting task. In this study, three retrieval tasks describing, searching and interacting were examined. The participants described their search topic, listed search keywords or phrases, and drew pictures which represented their topic in a pre-search questionnaire. Studying this questionnaire, the investigator collected data on the describing and searching tasks. During the search, the participants recorded the search keywords or phrases they actually used. They drew pictures, which they retrieved in a

post-search questionnaire. The investigator also collected data on the interacting tasks from the post-search questionnaire.

Jorgensen collected the user's verbal queries to study the user's retrieval tasks. Enser (1995) states that users can express their queries in the following four modes: LL, VV, VL, and LV. According to Enser's four modes, studying the user's query in both linguistic and visual modes would lead to a greater understanding of the user's image retrieval tasks.

In order to ascertain whether a user can generate a query in the linguistic and/or the visual mode when he/she is doing the three different tasks (describing, searching and interacting), an analysis of correlation coefficients is needed. For this purpose, the investigator studied the relationship between the topic description and keywords submitted by a user before and after his/her searches (the linguistic mode within the three retrieval tasks). For the visual mode, within the retrieval tasks, the investigator studied the number of drawings submitted by a user before and after his/her searches and the user's level of confidence in the drawings to see if there was any relationship between the searches. The investigator also studied the number of search keywords, the number of drawings and the level of confidence in the drawings submitted by the user before and after the searches to see if there was any relationship between the linguistic and visual modes.

Kuhlthau (1991), p. 364 suggests that an investigation of the search outcome in relation to cognitive and affective aspects of the information-seeking process is needed. Her study took place in six high schools with 147 seniors in English classes selected as participants. She studied the following variables:
- the student's degree of confidence during the information-seeking process,
- the focus of the student's paper as judged by the instructor, and
- the instructor's grade on the student's written paper.

In accord with Kuhlthau's findings, this study collected the user's judgement of the success of his/her search. The investigator examined the relationship between the degree of success as judged by the user and the instructor, and the other study factors: gender, experience with online public access catalogs (OPACs), experience with electronic databases, and the number of keywords and drawings submitted by the user before and after the search. The investigator interviewed the students to obtain reasons for their judgement on their degree of success, and the instructor for the grading criteria of the written paper. The investigator was interested in the following research questions:
1. Can any associations be identified between the three tasks (describing, searching and interacting) and query modes?
2. Is there a relationship between the user's characteristics (gender and IR experience), the user's judgement on the degree of success of his/her search and the grade on the written paper?

The following specific questions were addressed:
- Did the user's gender have any influence on the success of his/her retrieval tasks?
- Did the user's experience with OPACs and electronic databases affect his/her search behavior?
- Did the user use mostly keywords or phrases from his/her topic description?
- Was the user able to represent his/her search topic both in the linguistic and visual modes?
- When the user generated more textual queries, did he/she feel more confident about drawing a picture to represent his/her search topic?
- After an interacting task, was the user able to represent his/her search topic both in the linguistic and visual modes?
- Did the user finish his/her search successfully?

As Garber and Grunes (1992) point out, an art image retrieval system should allow a user to describe an image by entering explicit descriptions of the image (linguistic mode), or the user may target an identified image and use that as a source for finding similar images (visual mode). Based on this understanding, a user is expected to be able to express his/her image queries successfully either in the linguistic or in the visual mode when he/she is doing the three tasks (describing, searching, and interacting). However, relationships among the three tasks within the two modes need to be examined.

For example, when a user is able to generate many queries (a searching task) in the linguistic mode, is he/she confident enough to draw a typical picture to represent his/her search topic (a describing task in the visual mode)? If the answer is no, the relationship between the searching task in the linguistic mode and the describing task in the visual mode is not meaningful. On the other hand, the user may be confident about drawing a picture to represent his/her search topic, but he/she may not know how to extract some search keywords or phrases to search for images. If this is so, art image retrieval systems should provide both textual and visual information to help users expand the search spectrum.

A variety of parameters were used to examine relationships among the three tasks (describing, searching, and interacting) within the modes (linguistic and visual). Meanwhile, the user's gender, experience with OPACs and electronic databases, degree of success for the search results, and grade on the written paper were also studied.

The parameters were:
- gender,
- experience with OPACs,
- experience with electronic databases,
- the describing task (the percentage of keywords or phrases participants planned to use that were found in the topic title or topic description),
- the number of keywords or phrases participants planned to use,
- the number of drawings participants used to represent the topic title before the search,
- the level of confidence measured by a 7-point scale of participants in their drawing representing the topic title before the search,
- the number of keywords or phrases participants actually used in the search,
- the number of drawings participants used to represent the topic title after the search,
- the level of confidence on a 7-point scale of participants in their drawing representing the topic title after the search,
- the degree of success for the search results based on a 7-point scale ($1 =$ extremely unsuccessful to $7 =$ extremely successful) judged by participants themselves, and
- the grade on the written paper on a 4-point scale ($0 =$ failed to $4 =$ excellent) given by the instructor.

4. Research design

4.1. Background

This research study was conducted in the context of two medieval art courses offered by the Department of the History of Art and Architecture at the University of Pittsburgh during the

1998 Fall semester. The department has an important information resource, the Frick Fine Arts Library. This library is a non-circulating research collection serving scholarly needs for the Department of the History of Art and Architecture, and undergraduate needs in Fine Arts and Studio Arts courses. The collection contains over 75,000 volumes and maintains subscriptions to over 200 journals in relevant fields. Of special interest are in-depth holdings in the fields of medieval, renaissance, and modern art and the facsimile collections of medieval manuscripts and oriental scrolls. Another information source is the visual resources laboratory which is equipped with eight Apple Macintosh computers connected to the University campus network, two slide scanners, two flatbed scanners and one laser printer. Each computer has word processing, image processing and Web browser software. The lab is open to the students, faculty and staff of the Department only.

4.2. Participants

The two medieval art courses, HA & A 0240 Medieval Artistic Patronage and HA & A 1200 Special Topics in Medieval Art and Architecture: Medieval Cities, are undergraduate level courses for students majoring in the history of art and architecture, but graduate students may take the courses as a prerequisite for their graduate studies. HA & A 0240 is an introductory course and HA & A 1200 is an advanced course. The students in the two medieval art courses are required to finish a written paper to fulfill their course requirements. The instructor provides a list of topics within the themes of the classes. The instructor specifically requires the students to include at least twenty images related to their individual research topics in the paper, which must be in the context of medieval art and approved by the instructor. Images can be either in digital or hard-copy form, and image resources can be monographs in a library collection or a museum art collection, museum publications such as postcards and catalog, web sites on the Internet, or any other source. Students are encouraged to explore their searches in different environments.

4.3. 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. They also drew pictures to represent their search topic. The participants were required to fill out the questionnaire and 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 also asked the participants to draw pictures they retrieved. The investigator was interested in whether the participants were more confident in their post-search drawings compared to their pre-search drawings. The post-search questionnaire was collected with the written paper.

4.4. Post-search interview

Prior to the post-search interview, the investigator reviewed the questionnaires. Each participant was asked to explain ambiguous answers or points. The participants were also asked to fill

out their demographic information. Chen (1999) has described the detailed research methodologies and data analysis methods in his dissertation.

5. Research findings

5.1. Participants' characteristics

Eighteen students registered for HAA 0240 and 20 for HAA 1200. Due to several personal factors, some students withdrew from the two courses. Eleven students in the HAA 0240 class and 18 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 1 presents the participants' characteristics in terms of academic status, gender, age and information retrieval (IR) experience. IR experience was judged by the frequency of use of OPACs and electronic databases. Tables 2 and 3 show the distribution of number of search keywords/phrases and drawings submitted by the two courses of participants before and after their searches.

Table 1
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	6.9
	5. Masters	3	10.35
Gender	1. Male	13	44.83
	2. Female	16	55.17
Age	1. Under 20	2	6.9
	2. 20–25	20	68.97
	3. 26–30	3	10.35
	4. 31–35	1	3.45
	5. 36–40	2	6.9
	6. 41–45	0	0
	7. 46–50	0	0
	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	1. None	14	48.28
(other than OPACs)	2. Daily	1	3.45
	3. Weekly	0	0
	4. Monthly	5	17.24
	5. Less than monthly	9	31.04

Table 2
Distribution of number of search keywords/phrases

Course	Pre-search				Post-search			
	Min	Max	Mean	S.D.	Min	Max	Mean	S.D.
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 3
Distribution of number of drawings

Course	Pre-search				Post-search			
	Min	Max	Mean	S.D.	Min	Max	Mean	S.D.
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

Three participants changed their topics after completing the pre-search questionnaires. These were excluded from the analysis of task behavior and query modes.

5.2. Task behavior and query modes

The study parameters are participants' gender status, IR experience, the describing task, the searching task, the interacting task, the level of success, the grade on the written paper, the linguistic query mode and the visual mode.

Two-tailed Pearson correlation coefficients were used to test the relationships among the following parameters:
- gender,
- experience with OPACs,
- experience with electronic databases,
- the describing task (the percentage of keywords or phrases participants planned to use that were found in the topic title or topic description),
- Pre3: question 3 of the pre-search questionnaire (the number of keywords or phrases participants planned to use),
- Pre4: question 4 of the pre-search questionnaire (the number of drawings participants used to represent the topic title before the search),
- Pre5: question 5 of the pre-search questionnaire (the level of confidence, a 7-point scale, of participants in their drawing representing the topic title before the search),
- Post1: question 1 of the post-search questionnaire (the number of keywords or phrases participants actually used in the search),

- Post2: question 2 of the post-search questionnaire (the number of drawings participants used to represent the topic title after the search),
- Post3: question 3 of the post-search questionnaire (the level of confidence, a 7-point scale, of participants in their drawing representing the topic title after the search),
- success: the degree of success for the search results based on a 7-point scale (1 = extremely unsuccessful to 7 = extremely successful) judged by participants themselves, and
- grade: the grade on the written paper on a 4-point scale (0 = failed to 4 = excellent) given by the instructor.

Table 4 shows that the mean percentage of keywords or phrases the 26 participants planned to use that matched words or phrases in their topic title or topic description was only 33.66% (S.D. = 0.23). The mean number of keywords or phrases participants planned to use in the search was 7.00 (S.D. = 2.08), the mean number of drawings they used to represent the topic title before the search was 2.65 (S.D. = 1.60), and the mean of the level of confidence (on a 7-point scale, 1 = not confident at all to 7 = extremely confident) in these drawings representing the topic title was 2.90 (S.D. = 1.58). The mean number of keywords or phrases participants actually used in the search was 9.42 (S.D. = 4.72), the mean number of drawings participants used to represent the topic title after the search was 3.46 (S.D. = 1.20), and the mean of the level of confidence in drawings representing the topic title was 3.80 (S.D. = 1.62). The mean of the level of success was 5.73 on a 7-point scale (S.D. = 1.22) and the mean score of the grade was 3.36 (S.D. = 0.75) on a 4-point scale.

5.3. Gender as a concern

Of the 26 participants, 16 were female and 10 were male. The investigator was interested in whether the parameter of gender played an important role in image retrieval tasks. Two-tailed Pearson correlation coefficients were applied to the mean scores of the study parameters and the results showed that there was a significant difference between gender and experience with OPACs ($r = 0.42$ and $P = 0.03$). This indicated that female students in this sample had a higher user frequency of OPACs than male students. No significant differences were found between gender and the other study parameters.

Table 4
Distribution of study parameters for statistical analysis ($N = 26$)

	Minimum	Maximum	Mean	S.D.
Describing task	0.00%	80%	33.66%	0.23
Pre3	3.00	11.00	7.00	2.08
Pre4	0.00	6.00	2.65	1.60
Pre5	1.00	5.60	2.90	1.58
Post1	4.00	21.00	9.42	4.72
Post2	2.00	6.00	3.46	1.20
Post3	1.00	7.00	3.80	1.62
Level of success	4.00	7.00	5.73	1.22
Grade	0.00	4.00	3.36	0.75

5.4. Personal information infrastructure as a concern

Personal information infrastructure is an important factor in a user's information-seeking behavior. IR experience is part of a personal information infrastructure. In this study, IR experience was defined by two measures: experience with OPACs and experience with electronic databases. Most participants had experience with OPACs but less with electronic databases. Four participants (15.39%) used OPACs daily, seven (26.92%) weekly, 10 (38.46%) monthly and five (19.23%) less than monthly. Fourteen (53.85%) participants had some experience with electronic databases and 12 participants (46.15%) had not used any electronic databases at all. Two-tailed Pearson correlation coefficients were used to investigate the relationship between the participants' information infrastructure and image retrieval tasks.

5.4.1. Experience with OPACs as a concern

Two-tailed Pearson correlation coefficients showed that there was a significant difference between experience with OPACs and Post3 (the level of confidence for drawings representing the topic title after search) ($r = 0.37$ and $P = 0.06$). This indicated that those of the 26 participants who had more experience with OPACs felt more confident about their drawings representing their topic titles after their searches. There was also a significant difference between experience with OPACs and the degree of success for the search results ($r = 0.36$ and $P = 0.07$). Those of the 26 participants, who used OPACs more often had a higher degree of success for their search results.

5.4.2. Experience with electronic databases as a concern

There was no significant difference between experience with electronic databases and the other study parameters. This showed that experience (or lack thereof) with electronic databases did not play an important role in this study.

5.5. Relationships among the describing task, the searching task and interacting task

An important focus of this study was to discover relationships among the three different tasks: describing, searching and interacting. The following sections address the investigation of these relationships.

5.5.1. Relationship between the describing task and searching task

The mean percentage of the pre-search keywords or phrases drawn from the topic titles and subject descriptions was 33.66. There are two possible reasons for the findings:
- most participants did not know how to describe their information needs, and
- they did not know how to extract keywords from the topic and description.

The relationship between the describing and searching tasks was not found to be very strong in this study.

5.5.2. Relationship between the searching task and interacting task

T-tests, one-way ANOVA and Pearson correlation coefficients were applied to examine the following three pairs of parameters (Table 5):

Table 5
Relationship between the searching and interacting tasks

Parameter	T-test (t)	F-test (F)	Pearson (r)
Pre3 and Post1	$t = 2.96$, $P = 0.01$**	$F = 8.75$, $P = 0.01$**	$r = 0.465$, $P = 0.02$***
Pre4 and Post2	$t = 2.27$, $P = 0.03$***	$F = 5.17$, $P = 0.03$***	$r = 0.190$, $P = 0.353$
Pre5 and Post3	$t = 2.20$, $P = 0.04$***	$F = 4.83$, $P = 0.04$***	$r = 0.160$, $P = 0.436$

** Correlation is significant at the 0.05 level (2-tailed).
*** Correlation is significant at the 0.01 level (2-tailed).

- Pre3 (the number of keywords or phrases participants planned to use) and Post1 (the number of keywords or phrases participants actually used in the search),
- Pre4 (the number of drawings participants used to represent the topic title before the search) and Post2 (the number of drawings participants used to represent the topic title after the search), and
- Pre5 (the level of confidence for drawings representing the topic title before the search) and Post3, (the level of confidence for drawings representing the topic title after the search).

The results of a *t*-test ($t = 2.96$, $P = 0.01$) and one-way ANOVA ($F = 8.75$, $P = 0.01$) showed that the mean number of search keywords or phrases (9.42) participants actually used in the interacting task was significantly greater than the mean 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 actually use more keywords or phrases in the interacting task.

The results of a *t*-test ($t = 2.27$, $P = 0.03$) and one-way ANOVA ($F = 5.17$, $P = 0.03$) showed that on 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, the results of a *t*-test ($t = 2.20$, $P = 0.04$) and one-way ANOVA ($F = 4.82$, $P = 0.04$) showed that on average participants had a higher degree of confidence in drawings made after their searches (mean = 3.80) than before their searches (mean = 2.90). However, there was no significant relationship between the level of confidence in drawings before and after, indicating that individual students who had a relatively higher degree of confidence in their drawings before their searches did not necessarily have a relatively higher degree of confidence in their drawings after their searches.

Regarding the level of confidence, the 26 participants gave relatively low scores both on the searching (2.90) and interacting (3.80) tasks in the visual mode 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 confident at all about their drawings representing their topic title either before or after their searches.

5.5.3. Relationship between the describing task and interacting task

Pearson correlation coefficients showed a significant relationship between the describing task and the mean number of keywords or phrases participants actually used ($r = -0.45$ and $P = 0.02$). This indicated that the participants who drew fewer pre-search keywords from their topic title and topic description actually used more keywords or phrases during their searches, but it was not known if the additional keywords or phrases used during the search process were also drawn from the topic title and topic description. Those participants did explore more queries in the interacting task than the participants who drew more pre-search keywords/phrases from their topic title and topic description.

5.6. Relationship between the linguistic and visual modes

The investigator was also interested in the relationship between the linguistic and visual modes within the three different tasks (Table 6). Pearson correlation coefficients showed a significant relationship between the number of keywords or phrases participants planned to use and the number of pre-search drawings ($r = 0.43$ and $P = 0.03$) they made. This indicated that the participants who generated more keywords or phrases they planned to use also produced more drawings representing their topic title.

The same relationship was also found between the number of keywords or phrases participants actually used and the number of post-search drawings ($r = 0.42$ and $P = 0.03$) they made. This indicated that the participants who actually generated more keywords or phrases during their searches also produced more drawings representing their topic title.

Pearson correlation coefficients showed a significant relationship between the means of the number of pre-search drawings and the level of confidence in these drawings, while the same result could not be found between the means of post-search drawings and the level of confidence in these drawings. This indicates that during the interacting task the participants who produced more drawings were not necessarily more confident than participants who produced fewer drawings.

Table 6
Relationship between the linguistic and visual modes

Parameter	Pearson (r)
Describing task and Pre4	$r = -0.14$, $P = 0.49$
Describing task and Pre5	$r = 0.15$, $P = 0.46$
Pre3 and Pre4	$r = 0.43$, $P = 0.03^*$
Pre3 and Pre5	$r = 0.09$, $P = 0.67$
Pre4–Pre5	$r = 0.42$, $P = 0.03^*$
Post1–Post2	$r = 0.34$, $P = 0.09^{**}$
Post1–Post3	$r = 0.16$, $P = 0.42$
Post2–Post3	$r = 0.19$, $P = 0.42$

[*] Correlation is significant at the 0.05 level (2-tailed).
[**] Correlation is significant at the 0.10 level (2-tailed).

In the pre- and post-search questionnaires, 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 to have 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.

The participants were required to draw pictures to represent their topics but one participant did not draw any pictures because she could not find an image in her mind. This situation illustrated the importance of the participant's cognitive ability.

5.7. Relationship between the level of success and the three tasks

Another research question concerns the relationship between the level of success and the three tasks (describing, searching and interacting). Pearson correlation coefficients showed a significant relationship between the means of the level of success and the describing task ($r = -0.37$ and $P = 0.06$). This showed that the participants who had a lower percentage of keywords found in Questions 1, 2 and 3 of pre-search questionnaire gave a higher score for the level of success. The same result was found between the means of the number of pre-search drawings and the level of success ($r = -0.36$ and $P = 0.07$). The participants who had fewer pre-search drawings gave a higher score for the level of success.

In the post-search questionnaire, 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 20 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 the participants reported that the search process was not difficult.

5.8. Relationship between grade on the written paper and the other study parameters

The participants' level of success agreed with the grade on their written paper given by the instructor. Pearson correlation coefficients found a significant relationship between the means of the level of success and the grade on the written paper ($r = 0.33$ and $P = 0.10$). However, there was no significant relationship between the grade on the written paper and the other study parameters.

6. Discussion

The objective of this study was 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 judgement of the degree of success of his/her search and the grade on the written paper.

The following questions were addressed:

Did 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 the gender and the other study parameters.

Did the user's experience with OPACs and electronic databases affect his/her search behavior? Experience with OPACs was related to the participants' level of confidence in their post-search drawings and the level of success for the search results. No relationship was found between experience with electronic databases and the other study parameters.

Did the user use mostly keywords or phrases from his/her topic description? The findings of this study showed participants drew only 33.66% of the pre-search keywords or phrases which were drawn from their topic titles and topic descriptions. 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 that there was a significant relationship between the describing task and the number of search keywords or phrases participants planned to use ($r = -0.36$ and $P = 0.07$). 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.

Was the user able to represent his/her search topic both in the linguistic and visual modes? The findings of this study showed 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 in 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 showed 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 generated more textual queries, did he/she feel more confident about drawing a picture to represent his/her search topic? Pearson correlation coefficients showed a significant relationship between the number of keywords or phrases participants planned to use and the number of pre-search drawings ($r = 0.43$ and $P = 0.03$) participants made. This indicated that the participants who generated more keywords or phrases which they planned to use also sketched more drawings representing their topic title.

Pearson correlation coefficients showed a significant relationship between the number of keywords or phrases actually used and the number of post-search drawings ($r = 0.42$ and $P = 0.03$) participants made. This indicated that the participants who generated more keywords or phrases during their searches also produced more drawings representing their topic title.

Pearson correlation coefficients also showed a significant relationship between the mean number of pre-search drawings and the mean score of the level of confidence in these drawings, although the same result could not be found between the mean number of post-search drawings and the mean score of the level of confidence in 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 in their pre- and post-search drawings, the 26 participants gave themselves relatively low scores both on the searching (2.90) and interacting (3.80) tasks in the visual mode 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 confident at all about their drawings representing their topic title either before or after their searches.

The 26 participants reported their reasons for their degree of matching scores:
- 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 to be 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, was the user able to represent his/her search topic both in the linguistic and visual modes? The results of a *t*-test ($t = 2.96$, $P = 0.01$) and one-way ANOVA ($F = 8.75$, $P = 0.01$) showed that the number of search keywords or phrases (9.42) participants actually used in 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 in the interacting task.

The 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 produced 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, the results of a *t*-test ($t = 2.20$, $P = 0.04$) and one-way ANOVA ($F = 4.82$, $P = 0.04$) showed that, on the average, students had a higher degree of confidence for 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.

Did the user finish his/her search successfully? Overall, the participants who gave themselves a higher score on the level of success received a better grade on 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. Pearson correlation coefficients showed a significant relationship ($r = 0.33$ and $P = 0.10$) 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 participants (25, 96.15%) retrieved 20 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 the participants reported that 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 20 images covered the range of topics 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. 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.

7. Conclusions

The participants in this study, were not confident that their drawing skills were adequate to represent the images they were seeking. To educate future art historians successfully, the results of this study suggest that drawing instruction should be considered in the design of art history curriculum. 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 also be taken into account. 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. The ability to generate queries in a visual mode 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 circle, square or triangle to present an image query. The user can also generate a more complex sketch to present the search topic.

In this study, the results of a *t*-test showed that the participants drew more pictures after their searches and the mean score of the level of confidence was also higher after the search. However, the mean scores of the level of confidence in their drawings both before and after their searches were still relatively low (2.90, pre-search and 3.80, post-search, on a 7-point scale). This indicates that the participants did not feel confident about the drawings representing their topics, even after seeing some relevant images. The majority of the participants (21, 80.77%) did not consider themselves as having good drawing skills either before or after their searches. Some participants were not sure about the details of the topic and/or the details of the images before the search.

The significant relationship between the 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. However, the two components were not controlled in this study and more concrete evidence could not be obtained. For further studies, the factor of IR experience should be controlled to investigate the effect of library instruction programs on user's image retrieval tasks. The results may be used to improve students' information retrieval skills as part of their overall education.

The designers of content-based image retrieval systems could take advantage of the user studies to evaluate the usability of current search functions and input devices and to see whether these search functions and input devices provide necessary assistance to users. Would the user feel confident about constructing a visual query with these search functions/devices? Is any user training needed for the current search functions/devices? How much effort is required for a user to learn the search functions/devices? The results can be applied to refine the current search functions/device and to invent new search functions and devices.

Based on the findings of this study, the investigator is interested in the following research topics for his future studies:

* *Usability of current search functions of content-based image retrieval systems.* The effect of search functions/devices of the image retrieval systems on the user search behavior should be considered for further study. 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.
* *Cognitive models.* No relationships were found between the linguistic mode and visual mode of queries in this study. The participants also did not show a high level of confidence in drawings representing their topics both before and after searches. It is suggested that further studies on cognitive models should be considered to discover relationships between the linguistic mode and the visual mode, and to further understand the user's ability to express ideas in a visual format.

Acknowledgements

References

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

Allen, B. L. (1996). *Information tasks: Toward a user-centered approach information systems*. New York: Academic Press.

Bates, M. J., Wilde, D. N., & Siegfried, S. (1993). An analysis of search characteristics: Beyond the topical approach to information retrieval. *Library Quarterly*, *63*, 1–39.

Beard, D. V. (1991). Computer human interaction for image information systems. *Journal of the American Society for Information Science*, *42*(8), 600–608.

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.

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

Case, M. (1996). Document for dialogue: Categories for the description of work of art. *Visual Resources*, *11*, 257–270.

Cawkell, A. E. (1994). *A guide to image processing and picture management*. London: Gower Press.

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*, 12(40), 63–71.

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

Chen, H., & Rasmussen, E. M. (1999). Intellectual access to images. *Library Trends*, *48*(2), 291–302.

Dervin, B., & Nilan, M. (1986). Information needs and uses. *Annual Review of Information Science and Technology*, *21*, 3–33.

Dervin, B. (1989). Users as research inventions: how research categories perpetuate inequities. *Journal of Communication*, *39*(3), 216–232.

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.

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

Gupta, A., Santini, S., & Jain, R. (1997). In search of information in visual media. *Communications of the ACM*, *40*(12), 35–42.

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

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

Heidorn, P. B. (1997). Natural language processing of visual language for image storage and retrieval. Doctoral dissertation. University of Pittsburgh, unpublished.

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 and 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). Doctoral dissertation. Syracuse University, unpublished.

Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science*, *42*(5), 361–371.

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

Layne, S. S. (1997). Modeling relevance in art history: Identifying attributes that determine the relevance of art works, images, and primary text to art history research. Doctoral dissertation. University of California, Los Angeles, unpublished.

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.

Stam, D. C. (1989). Tracking art historian: On information needs and information-seeking behaviour. *Art Libraries Journal*, *14*(3), 13–16.

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.

Wiberley, S. E. (1988). Names in space and time: The indexing vocabulary of the humanities. *Library Quarterly*, *58*, 1–28.