Visual Access for Retrieving Emotional Content of Pictures

Jee Yeon Lee
School of Information Studies
Syracuse University
Syracuse, NY, U.S.A.

Abstract *The traditional and still pervasive way of access to image information is through descriptive keyword searching. However, many studies in image indexing and retrieval have reached a consensus on the difficulties and limitations of text-based image description. This situation is more obvious when the description involves subjective interpretation of the content of images. One of the best examples that depend on relatively high degree of subjective interpretation is emotion. This research investigates the feasibility of using visual browsing, that is being used comparatively much less than keyword searching, for the retrieval of emotional content of pictures. The effectiveness of keyword access versus visual access were examined through experiments in which participants searched for pictures of specified emotions using different access modes: keywords only, visual browsing only, and the combination of both. Findings and results can guide design of image retrieval systems, especially the retrieval system of subjective and interpretive information.*

Keywords: Image Retrieval, Access Modes, Visual Browsing

1 Introduction

Access to image information is achieved through largely two different ways: text-based access and content-based access. The former relies the mechanism of image retrieval on the textual description of images and use of matching keywords ([1]; [2]; [3]; [4]; [5]). The content-based image retrieval is based on the visual features such as shape, color or texture ([6]; [7]). However, none of these two major trends of image retrieval is not yet regarded as an appropriate and sufficient way of dealing with the complexity of information delivered by the content of images. Textual description of images can never be perfect in terms of its universality and objectivity. Content-based image retrieval is still in its infancy in terms of the types of information that can be extracted - factual and tangible.

Starting from the consideration of the difficulties of text-based image description and the unripe nature of content-based image access, this research explores a relatively new approach, the use of visual browsing technique enabling users to directly react to what they see and feel while viewing images. The visual browsing technique used in this research does not mean "random looking" until a sufficient number or precisely matching image is found but means selecting an image through "looking" and then following some "browsing path" to related images. The browsing path was established on the basis of descriptors assigned to images. Strictly speaking, the visual browsing technique used in this research is a text-based visual access mode as an alternative to keyword access.

In addition to the use of comparatively new access mode, visual browsing, there is another central issue of this research: emotional content of pictures. Difficulties of text-based image description are more obvious and serious when the description deals with the secondary subject matter rather than the primary subject matter ([8]; [9]). Emotion is one of the best

examples of the secondary subject matter which is based on an interpretation of objects seen in pictures. Goodrum also points out, by mentioning Keister's analysis of query logs, that a user's emotional needs are highly subjective and "do not lend themselves to indexing" ([10], p.65). This study deals with emotions, not elicited by pictures, but portrayed and revealed by the content of pictures. Emotional content can be a difficult feature for the retrieval of image information because emotional information relies more on the human judgment to interpret it. It is unrealistic to expect that the interpretation of emotional information by indexers always matches the interpretation by viewers or users of image collections.

2 Background

Pictures are rich in information but pictures are also complex and equivocal information sources. A single picture may meet a variety of information needs, but all the different visual factors which meet each different information need cannot always be articulated by an observer because of the different forms of the expression (in text form) and presentation of information (in visual form). Word-based articulation is limited in describing subtle nuances of visual presentation of human sense.

Two photos, illustrated in Figure 1 and Figure 2, are the examples of the diversity, subjectivity, and condition-dependence of interpretation made by participants of the researcher's preliminary study. The two photos show the difficulties in describing and indexing images. People may agree that those pictures convey strong emotional information, but emotions they read from the pictures may vary. Indeed, findings from the preliminary study on how much people agreed or disagreed on the emotional content of pictures showed that those pictures were read very differently among respondents. For example, the emotions read by eight respondents from the example photo in Figure 1 included happiness, amusement, cheering, pride, joy, excitement, satisfaction, surprise, and victory. One person even read the emotion of caring and compassion from the picture. Responses from the example photo in Figure 2 showed a similar pattern. Respondents read several different emotions, such as happiness, amusement, cheering, excitement, satisfaction, victory, and caring again. It is hard to expect that indexing words describe all the variety of subjective determination and/or interpretation made by different people.

Figure 1: Photograph 1

Figure 2: Photograph 2

It seems obvious that having information in a visual mode and retrieving it by a textual mode can lead to disagreements and misunderstanding by searchers. Wirshbo brings up an interesting question that whether emotions can be determined from words. She starts her discussion with an old and perplexing question, "which comes first, thoughts or words?" ([11], p.287). She states that words often produce contradictory meanings and ambiguous utterances, due to misunderstanding, conscious misapplication or creative revitalizing.

Using different access modes may be a solution to improving the retrieval of emotional content of pictures. The visual browsing method

provides a user with the capability of directly reacting to what he or she feels and expressing his or her next information request by simply selecting a picture which best represents his or her needs. The user does not need to pass through the step of translating his or her emotional perception into a textual explanation. The translation may filter a portion of what the user feels. The translation may not express exactly the same perception of the user. The translation can be a similar but slightly different idea from what the user actually feels.

3 Research Objectives

This research explores the issues related to the retrieval of emotional content of pictures in the context of the modes of access. The main purposes of this research are first, to see how well descriptive keyword searching works for the retrieval of emotional content of pictures and its limitations due to indexing problems, and second, to investigate ways of improving access to visual emotional information by an alternative to keyword access, visual browsing. What is the most effective access mode for the retrieval of emotions portrayed in pictures? Is one method better than the other (i.e., keyword retrieval vs. visual browsing)? Or, is the combination of the two methods better than a single method? There is little research to answer the questions. Therefore, this research serves as a substrate to increase our knowledge in the area of how different access modes can help in the retrieval of emotional content of pictures.

Capabilities and potentials of using different modes of searching are examined through seeking answers to the following research questions:

- Is the visual browsing technique effective in terms of improving access to the emotional content of pictures?
- What are strengths and drawbacks in using the different access modes?

4 Methodology

4.1 Research Approach

This research begins without any previously known characteristics or variables that are attributable to understanding the use of different searching modes to retrieve visual emotional information or, more generally, abstract and intangible qualities of images. This research begins with no assumption or hypothesis because the research has not found any previous study that can inform about the use of different searching modes for the retrieval of visual emotional information. This study is an attempt to elicit fundamental information about the characteristics associated with the use of different modes of searching for the emotional content of pictures, as well as an attempt to reveal unknown variable that can help designers of image retrieval systems to understand the effect of using different modes of searching for visual emotional information. Given the above conditions, it is clear that this study is the first of its kind; therefore, it is an exploratory study.

4.2 Research Design

A prototype system of image retrieval and an image database containing about 250 pictures were constructed and are available at http://elmer.syr.edu/~jylee/image. The pictures were indexed collaboratively by six information scientists. The indexing task was guided by the empirical data obtained from the researcher's preliminary study, investigating emotions frequently observed and agreed by viewers.

The most important consideration in deciding the relationship among the pictures for visual browsing was the relationship between textual indexing and the picture it describes. The keywords assigned to each picture were used to determine what other pictures are associated or close to it for visual browsing. For example, a picture with the indexing words of happiness, romance, smile, and kissing is associated with the pictures with at least one of the words. If a user clicks on the picture, the next screen displays all the pictures that have at least one of the words as their own indexing word(s). The order of

displaying the retrieved pictures is by the frequency of the keywords shared with the selected picture.

Twelve participants recruited from the field of information science, art, psychology, education, communication, and management, participated in the experiment. They were asked to search for pictures that portrayed three specified emotions by using different modes of searching (e.g., caring pictures by keywords only, sadness pictures by browsing only, and romance pictures by both methods). Emotions, modes of searching, and the order among the three access modes treated were rotated equally. While searching, the participants were asked to determine the relevance of the pictures retrieved.

Once the searching task was completed, each participant was asked to answer to the questions in a post-search questionnaire. The questionnaire asked the participants about ease of use, participants' preference of a method to the others and reason(s).

5 Results

This section reports results and findings from an ongoing study of access modes for the retrieval of emotional content of pictures. The study plans to gather more data to test the different modes of searching with all twenty-five emotions observed frequently in pictures by participants of the researcher's preliminary study. Findings reported here are based on the data set available at present.

5.1 Findings from Numeric Data

There are three broad categories of numeric data analyzed: precision, the number of relevant pictures retrieved, and the number of retrieved pictures browsed.

Precision

The participants made relevance judgment of the retrieved pictures while conducting the experiment. Table 1 shows the precision based on the relevance judgment.

Table 1: Precision measurement by each emotion

	Precision (%)			
Emotion	KO*	VBO	COB	Avg
Happiness (55**)	94.5	86.7	93.4	91.5
Excitement (47)	49.2	48.9	100.0	66.0
Sadness (45)	88.4	78.0	88.8	85.1
Grief (39)	59.1	77.1	85.7	74.0
Caring (32)	71.9	100.0	62.8	78.2
Defiance (29)	72.4	48.3	68.8	63.2
Pride (11)	90.9	80.0	63.9	78.3
Surprise (7)	61.9	31.7	32.3	42.0
Romance (6)	37.2	38.5	58.9	44.9
Lack of comm.(5)	35.6	52.3	41.1	43.0
Compassion (3)	90.0	6.7	82.5	59.7
Shyness (2)	25.0	10.0	21.1	18.7
Average	64.7	54.9	65.5	61.7

* KO = Keywords Only,
VBO = Visual Browsing Only,
COB = Combination of Both, and
Avg = Average
** Occurrence

As seen in the table, the effectiveness of each different mode measured by precision is that using both modes results in the most successful search, keyword searching is the next successful one, and visual browsing only is the least successful search. It can also be seen that the difference of the average precision between keywords only and combination of both is not large, compared with the difference between these two methods and visual browsing only. The finding that the individual precisions are spread widely, from 0.0 to 100.0, might be caused by participants' subjective determination as well as the use of different emotions.

Another interesting finding on precision by different emotions is that the emotions represented by many pictures in the database collection tended to result in higher precision

than the emotions represented by a few pictures, except the emotion compassion. The participants who searched for the pictures of compassion found additional good pictures when they viewed the pictures of caring. Although the precision of compassion pictures by keywords only mode is higher than the other modes, the number of relevant pictures retrieved for the emotion is higher when they used the visual browsing or the combination mode than when they used the keywords only mode.

Number of Relevant Pictures Retrieved

A noteworthy point observed throughout the experiment is that the number of good pictures retrieved might be a good indicator for measuring the effectiveness of different access modes. Table 2 presents the data on the number of good pictures retrieved by different access modes.

Table 2: Number of Relevant Pictures

	Number of Relevant Pictures Retrieved		
Emotion	KO*	VBO	COB
Caring	23	15	38
Compassion	15	4	16
Defiance	21	27	30
Excitement	24	93	39
Grief	31	34	24
Happiness	52	32	28
Lack of comm.	5	20	9
Pride	20	12	40
Romance	7	37	32
Sadness	40	46	53
Shyness	2	3	3
Surprise	6	19	35
Total	246	342	347

* KO = Keywords Only,
VBO = Visual Browsing Only, and
COB = Combination of Both

The most effective mode is the combination mode, the next effective mode is the visual browsing only mode, and the least effective mode is the keywords only mode. The result implies that the majority of relevant pictures retrieved by the combination mode seemed to be obtained through visual browsing. Therefore, the result also implies that the visual browsing method seems to be more effective in terms of retrieving more relevant pictures than keyword searching if only those two methods are compared.

Judging from individual emotion and comparing only keyword searching and visual browsing, the study found that visual browsing retrieved more relevant pictures for eight emotions out of twelve (66.7%), keyword searching was better for three emotions (25%), and both were equal for one emotion (8.3%).

The reasons why visual browsing is better in browsing fast and retrieving many more good pictures may be because 1) the output screen which displays retrieved pictures without textual description of each pictures may help users in viewing pictures rapidly without any interruption or dispersion of attention to reading the textual description and 2) many more pictures retrieved and available for relevance judgment by the visual browsing technique provide users with better chance of retrieving many good pictures.

Number of Pictures Browsed

The number of pictures browsed by the participants is another good indicator for measuring the effectiveness of each different access mode. It is found that the participants tended to browse pictures rapidly while browsing visually than searching by keywords. Table 3 shows that how many pictures were viewed and used for relevance judgment by each participant.

Table 3 indicates that the participants using the visual browsing only mode browsed more than twice as many pictures as those using the keyword only mode. Ten participants (83.3%) viewed more pictures when using the visual browsing only mode than when using keywords only mode. The result regarding the number of pictures browsed by each emotion reveals a similar tendency. Participants browsed more using the visual browsing mode than using the keywords only mode. For nine emotions (75%), the participants viewed more pictures by visual

browsing than by keyword searching. The repeated measures analysis of variance found a significant difference between keyword access and visual browsing (df = 1, f = 9.883, sig. = .009, α = .05) as well as between keyword access and the combined mode (df = 1, f = 5.690, sig. = .036, α = .05). This finding indicates that the participants using the visual browsing mode, looked through the retrieved pictures rapidly, thus having a better chance to find more good pictures.

Table 3: Number of Pictures Browsed

	Number of Pictures Browsed			
PN*	KO	VBO	COB	Total
1	55	190	59	304
2	53	66	30	149
3	45	90	35	170
4	25	60	32	117
5	18	30	30	78
6	18	36	45	99
7	29	15	28	72
8	22	84	65	171
9	25	15	44	84
10	19	60	80	159
11	37	108	133	278
12	15	46	52	113
Total	361	800	633	1,794

* PN = Participant Number,
KO = Keywords Only,
VBO = Visual Browsing Only,
COB = Combination of Both

5.2 Findings from Non-numeric Data

The data obtained by the think-aloud method and the post-search survey were analyzed by the content analysis technique. The data were analyzed inductively because no initial analysis framework could be produced. This was due to the unavailability of previous research findings. However, some elements initially assumed by the researcher as potential popular answers, such as reasons for preferring a particular mode (e.g., ease of use, previous experience, etc.) guided data reduction and coding.

The results of analysis show that:

- The reasons why some participants (25%, n=12) preferred the keywords only mode to the other modes are

  - because the search terms that they wanted to use were available in the keyword list,
  - because it was easy to use emotion-related search terms, and
  - because they could make specific search requests using words.

- The reasons why other participants (16.7%, n=12) preferred the visual browsing only mode to the other modes are

  - because it helped in finding relevant pictures,
  - because they could browse pictures speedily, therefore, browsed more pictures,
  - because visual browsing provided the idea of what kind of pictures were available in the database, and
  - because visual browsing allowed to make the comparison of pictures visually.

- The reasons why the other participants (58.3%, n=12) preferred the combined mode to the other modes are

  - because they could take advantage of both keyword access and visual browsing,
  - because they could compare the pictures retrieved by the keyword mode and the visual browsing mode in order to have the best search set,
  - because the combined mode allowed to move to the other mode freely, and
  - because using both modes helped in retrieving more relevant pictures.

- The reasons why some participants (58.3%, n=12) liked the keywords only mode the least are

  - because the search was limited to the keywords assigned by indexers,

- because text does not sufficiently express what people read from visual information sources, and
- because searching by keywords was monotonous and boring.

- The reasons why other participants (33.3%, n=12) liked the visual browsing only mode the least are

  - because visual browsing has a possibility of getting lost and may be get off the right track,
  - because visual browsing is rather a novel, new, and unfamiliar way of searching,
  - because the result of search by visual browsing was not satisfactory, and
  - because some of the representative pictures were hard to decide upon.

- The reason why a participant (8.3%, n=12) liked the combined mode the least is because the combined mode retrieved irrelevant pictures and led to a wrong direction in the search.

6 Discussion

Findings and results found positive potential for the visual browsing technique to improve the current state of image retrieval, especially the retrieval of subjective and interpretive visual information. There are several points that reveal the strength of the visual browsing method, used as alone or combined with the keyword access mode. First, the visual browsing mode retrieved more relevant pictures than keyword searching. The average precision of visual browsing (54.9%) is lower than that of keyword access (64.7%), but that is due to the fact that people looked through more pictures while using visual browsing, which made them encounter more irrelevant pictures as well as more relevant pictures. Indeed, some participants expressed that they favored visual browsing compared with keyword access because they could scan pictures speedily and find more good pictures.

Second, the visual browsing technique worked better than the keyword mode in terms of providing an overview of the database collection. Visual access provide users with the feature (i.e., providing an overview of the database collection) when they browse an initial screen that displays representative pictures. Users can have an idea of what kind of pictures are stored in a database. The feature is believed as useful for developing a search strategy/plan for the following steps.

Third, visual browsing enables people to be free from the burden that they should translate their information needs (i.e., needs for visual emotional information) into a textual explanation. Users of visual browsing can interact with the system visually. One of the major reasons why visual browsing was preferred by several participants was because they could make the comparison of pictures visually.

Fourth, visual browsing provides more access points to retrieve pictures. In keyword searching, as Goodrum notices from her study of image queries, most participants (66.7%, n=12) used just one or two search term(s), including the emotion word given from a search query ([12]). The participants used various sorts of visual clues, such as different styles of facial expression, body postures, action, and contextual information, which helped them to select and click on a picture to retrieve other similar pictures. When performing a search visually, they detected a variety of visual clues and used them as access points to retrieve other good pictures.

7 Further Research

This research can be viewed as a very first step toward investigating the feasibility of using visual browsing for the retrieval of emotional content of pictures. This paper reported preliminary results of this research, such as measuring the effectiveness of keyword access versus visual access and pros and cons of using different access modes. Data gathering will continue. Microscopic analyses of the study issues, such as similarities or differences of reactions to the different access modes by user types (e.g., gender, age, race, and educational level), visual clues useful to identify a specific emotion, and participants' mental reasoning to move to the next step, will be conducted with more data set available in the near future.

From a macroscopic view, the finding that visual browsing has positive potential for enhancing the current state of image retrieval calls future research that investigates whether the effectiveness of visual browsing can be generalized in all types of image information, including factual and objective information, or other intangible visual information, such as social message/notion, etc.

References

[1] Zheng, M. Metadata Elements for Object Description and Representation: A Case Report from a Digitized Historical Fashion Collection Project. *Journal of the American Society for Information Science*. 50(13): 1193-1208, 1999.

[2] Rasmussen E. Indexing Images. *Annual Review of Information Science and Technology*, 32: 169-196, 1997.

[3] Hastings, S. Query Categories in a Study of Intellectual Access to Digitized Art Images. *Proceedings of the Annual Meeting of the American Society for Information Science*, 32: 3-8, 1995.

[4] Keister, L. User Types and Queries: Impact on Image Access Systems. In *Challenges in Indexing Electronic Text and Images* (Fidel, R. et al., Eds.) Medford NJ: Learned Information, Inc., 1994.

[5] Enser, P. Query Analysis in a Visual Information Retrieval Context. *Journal of Document and Text Management*, 1(1): 25-52, 1993.

[6] Idris, F. & Panchanathan, S. Review of Image and Video Indexing Techniques. *Journal of Visual Communication and Image Representation* 8(2): 146-166, 1997.

[7] Aigrain, P. et al. Content-Based Representation and Retrieval of Visual Media: A State-of-the-Art Review. *Multimedia Tools and Applications* 3(3): 179-202, 1996.

[8] Markey, K. Access to Iconographical Research Collections. *Library Trends*, 37(2): 154-174, 1988.

[9] Leung et al. Picture Retrieval by Content Description. *Journal of Information Science*, 18: 111-119, 1992.

[10] Goodrum, A. Image Information Retrieval: An Overview of Current Research. *Journal of Informing Science* 3(2): 63-67, 2000.

[11] Wirshbo, E. Can Emotions Be Determined from Words? A Consideration of Recent Military Usage. *American Behavioral Scientist*, 33(3): 287-295, 1990.

[12] Goodrum A. & Spink, A. Visual Information Seeking: A Study of Image Queries on the World Wide Web. *Proceedings the 1999 Annual Meeting of the American Society for Information Science*, 36: 665-674, 1999.