A Comparative Study of Methods to Explore Searchers' Affective Perceptions of Images

Authors: JungWon Yoon

Introduction. Previous search query analysis studies demonstrated that emotional perceptions and needs tend not to be sufficiently reflected in image queries. To provide a better understanding of the emotional attributes of an image, this study explored and compared such attributes obtained from related tasks.

Method. This study analysed attributes of an image through sorting, describing and searching. A data-set was obtained from Flickr images tagged with one of six basic emotions (love, joy, surprise, anger, sadness and fear); fifty-nine students participated.

Results. The sorting and describing tasks revealed two categories of emotions, joy and sadness, which people perceive from images. They also revealed some emotional terms which were chosen frequently as labels for sorted images. Comparison of image perceptions obtained from the three tasks performed on an image demonstrated similar perceptions, but there are minor differences resulting from the unique characteristics of each task or research setting.

Conclusions. The categories of joy and sadness, along with the identified frequently used emotional terms, should receive attention in designing image indexing and retrieval systems. Also, the features of each task and environment setting should be considered when developing a research design exploring emotional aspects of an image.

Scholar Commons Citation
Yoon, JungWon, 'A Comparative Study of Methods to Explore Searchers' Affective Perceptions of Images' (2011). School of Information Faculty Publications. 127.
http://scholarcommons.usf.edu/si_facpub/127
A Comparative Study of Methods to Explore Searchers' Affective Perceptions of Images

Complete Abstract:

**Introduction.** Previous search query analysis studies demonstrated that emotional perceptions and needs tend not to be sufficiently reflected in image queries. To provide a better understanding of the emotional attributes of an image, this study explored and compared such attributes obtained from related tasks.

**Method.** This study analysed attributes of an image through sorting, describing and searching. A data-set was obtained from Flickr images tagged with one of six basic emotions (love, joy, surprise, anger, sadness and fear); fifty-nine students participated.

**Results.** The sorting and describing tasks revealed two categories of emotions, joy and sadness, which people perceive from images. They also revealed some emotional terms which were chosen frequently as labels for sorted images. Comparison of image perceptions obtained from the three tasks performed on an image demonstrated similar perceptions, but there are minor differences resulting from the unique characteristics of each task or research setting.

**Conclusions.** The categories of joy and sadness, along with the identified frequently used emotional terms, should receive attention in designing image indexing and retrieval systems. Also, the features of each task and environment setting should be considered when developing a research design exploring emotional aspects of an image.

This article is available at Scholar Commons: [http://scholarcommons.usf.edu/si_facpub/127](http://scholarcommons.usf.edu/si_facpub/127)
A comparative study of methods to explore searchers' affective perceptions of images

JungWon Yoon
School of Information, University of South Florida, Tampa, Florida

Abstract

Introduction. Previous search query analysis studies demonstrated that emotional perceptions and needs tend not to be sufficiently reflected in image queries. To provide a better understanding of the emotional attributes of an image, this study explored and compared such attributes obtained from related tasks.

Method. This study analysed attributes of an image through sorting, describing and searching. A data-set was obtained from Flickr images tagged with one of six basic emotions (love, joy, surprise, anger, sadness and fear); fifty-nine students participated.

Results. The sorting and describing tasks revealed two categories of emotions, joy and sadness, which people perceive from images. They also revealed some emotional terms which were chosen frequently as labels for sorted images. Comparison of image perceptions obtained from the three tasks performed on an image demonstrated similar perceptions, but there are minor differences resulting from the unique characteristics of each task or research setting.

Conclusions. The categories of joy and sadness, along with the identified frequently used emotional terms, should receive attention in designing image indexing and retrieval systems. Also, the features of each task and environment setting should be considered when developing a research design exploring emotional aspects of an image.

Introduction

The main purpose of information retrieval is to provide search results which can fulfill users' information needs. Therefore, it is critical to reflect users' information needs and their searching patterns in designing an information retrieval/indexing mechanism. Through search queries, users express their information needs and communicate with an information retrieval system. So, examining users' queries has been used as a way of understanding users' needs. In the image retrieval field, there have been several studies which analysed search queries under the same assumption (Armitage and Enser 1997; Chen 2001; Choi and Rasmussen 2003; Collins 1998; Enser and McGregor 1992; Goodrum and Spink 2001; Keister 1994). However, even with the benefits of studying query analysis, there are some issues that cannot be resolved by this method. For instance, search queries may not represent users' needs
completely, because users often do not recognize their own needs or have difficulty converting visual needs into verbal expressions. Also, experienced users of an image retrieval system tend to create queries which seem to return their anticipated search results, even if the users know that the queries do not represent their visual needs precisely. Therefore, in addition to query analysis, for understanding image searchers' unexpressed needs and image perceptions, several image related tasks, such as describing, sorting, similarity judgment and others, have been employed in image user studies.

Compared to query analysis study, studies exploring image related tasks consistently emphasized the significance of emotional attributes of an image. For instance, sorting task studies supported that people categorize images using abstract and emotional features of images (Jörgensen 1995; Rorissa and Hastings 1994; Laine-Hernandez and Westman 2006) and description task studies demonstrated that people tend to use more abstract and emotional terms in the describing tasks than in the search tasks (Jörgensen 1995; O'Connor, et al. 1999; Greisdorf and O'Connor 2002; Hollink, et al. 2004). A line of reasoning adopted by these studies explains that, because of the notion that emotional attributes are subject to an individual viewer's interpretation (Greisdorf and O'Connor 2002), current image retrieval indexing mechanisms have limitations in supporting access to subjective emotional meanings; therefore, searchers experienced in image retrieval systems are inclined not to use emotional terms in their queries (Jörgensen 2003; Eakins, et al. 2004).

In short, it has been demonstrated that emotional meanings offer significant messages, in which image users are interested. Therefore, in order to improve image retrieval effectiveness through emotional attributes, it would be necessary to examine those attributes of an image comprehensively. However, those previous studies explored emotional attributes only as a part of their studies; as a result, there was no overarching study comparing different research designs. We assume that by exploring and comparing emotional attributes obtained from different image related tasks, it will be possible to provide a better understanding of the emotional attributes of an image. Furthermore, comparing emotional features obtained from different tasks could affect how researchers design studies of emotional perceptions of images. As a more practical benefit, results of this study will suggest how emotional features can be integrated into image indexing and retrieval systems.

This study has the following two specific research questions:

1. How is an emotion-space (i.e., hierarical representation of emotions) represented through sorting and describing tasks? Are there differences in an emotion-spaces represented through sorting and describing tasks?
2. Are there differences in emotional perception for an individual image (i.e. emotions apprehended from each individual image) between searching, describing and sorting tasks?

**Literature review**

**Image related tasks**

If we could observe how people perceive, recognize and interpret visual materials, how they represent their needs for visual materials and how they operate their search processes, we would have a solid foundation for designing an image retrieval system or image organization scheme. However, since direct observation is not possible, researchers have attempted to understand human visual perception and searching activities by using several image related tasks. In this section, studies that examined image related tasks with the purpose of improving image retrieval performance are reviewed by focusing on key findings as well as research design. One image related task which is often incorporated in studies is describing. This task asks participants to describe what they notice in an image. Studies adopting describing were conducted on the basis of a relatively simple assumption; that is, attributes that are described while viewing an image reveal perceived or pertinent attributes of the image (Jörgensen 1995). However, since verbal description, which is not an inherent element of visual materials, has limitations in representing perceived or pertinent attributes, different settings have been adopted for exploring image perception from different perspectives. For example, Jörgensen (1995) used different types of descriptive tasks; O'Connor, et al. (1999) emphasized reactive description; Greisdorf and O'Connor (2002) used pre-defined queries; Laine-Hernandez and Westman (2006) compared keyword and free description. Comparison and implications of different approaches will be discussed later.
Another method to elicit perceived image attributes is sorting or categorization. Categorization has been a key topic in psychology, because it is how people process and make manageable an infinite number of different stimuli. In other words, people perceive the world and environment by categorization which treats non-identical stimuli as equivalent (Rosch 1977). Categorization is applied to information retrieval as a way to 'reduce search space and, thus, search time, because there are fewer categories than the total number of individual numbers' (O'Connor, et al. 1999: 681). Importantly, categorization can be used to facilitate browsing, which is useful when searchers have no idea what they are looking for or have difficulties in expressing their needs (Rorissa and Hastings 2004). Jörgensen (1995) assumed that categorization can reveal image attributes that are not extracted from describing tasks, because people sort non-identical images with a notion of similarity in which individual, cultural and contextual backgrounds are embedded. In addition, she suggested that the describing tasks in which participants focused on each individual image, closely mirrored known-item searching, whereas the sorting task in which participants looked at the whole set of images mirrored browsing. As will be discussed later, several studies demonstrated different distributions of image attributes between describing and sorting.

Collocating similar information objects (or documents) is a fundamental function of information organization; therefore, a good indexing mechanism should be able to collocate similar documents and differentiate dissimilar documents. In the field of image retrieval, researchers have attempted to evaluate their indexing mechanism by comparing it with similarity judgment tasks conducted by people (Rogowitz, et al. 1998; Rorissa 2005). Since the sorting task is also based on a notion of similarity, sorting and similarity judgment tasks have common features. However, whereas a sorting task usually involves participants sorting a set of images, similarity judgment tasks ask participants to compare a pair of images. Rogowitz, et al. (1998) found these two tasks demonstrated consistent results.

Related studies

Table 1 summarizes previous studies in terms of study purpose, image collection, participants, and procedure. There are a couple of studies which analysed different types of user descriptions. O'Connor, et al. (1999), who recognized the potential of users' reactions to images as access points, collected three types of users' reactions: caption, subject description and reactions (words or phrases describing how images makes users feel). The authors found the prevalent occurrences of narrative (conversational) and adjectival descriptors. Greisdorf and O'Connor (2002) investigated how users assign predetermined query terms to images and suggested the importance of affective and emotional query terms in image retrieval.

Several studies compared describing tasks with other image related tasks, such as search, sorting and similarity judgment. With the purpose of identifying typical image attributes which are revealed in image related tasks, Jörgensen (1995) conducted a set of describing tasks including descriptive viewing, descriptive searching and descriptive memory tasks, a sorting task and a concept search task. Content analysis on the set of describing tasks showed forty-seven attribute categories and twelve broad classes which can be characterized as either perceptual or interpretive attributes. When analysed by twelve classes, four perceptual classes, objects, people, colour and spatial location, were generally identified across all describing tasks, whereas emotion and abstract attributes appeared in the top half of the viewing task, they dropped to the midpoint for the descriptive search task. Jörgensen interpreted this result as the tendency for people who are formulating their search statements to do so with the realization that it will be processed by an image retrieval system.

Compared to describing tasks, the sorting task demonstrated a lower occurrence of the object, colour, spatial location and visual element classes and a higher occurrence of the abstract attribute. Also, a strong effect of sorting, based on whether or not a human was in a picture, was detected in the sorting task. The research concluded that compared to the describing tasks, the sorting task relied more on an interpretative attribute. Laine-Hernandez and Westman (2006) conducted a description (keywording and free description) and categorization study (hereafter, the categorization task is named sorting task, for consistent naming with other studies) to examine whether a particular image genre, journalistic photographs, influenced image indexing schemes. Location, people and descriptive terms were more often used in the free description task, whereas abstract concepts, themes, settings and emotions were more frequently used in the keywording tasks. Laine-Hernandez and Westman (2006) suggested that participants use more summarizing terms for the keywording tasks, because they had a limit of five terms. Overall, interpretational semantics were dominant in the describing tasks. In the sorting task, abstract and emotional themes and the presence of people were the main criteria in judging image similarity.
Hollink, et al. (2004) proposed and tested a framework for the classification of image attributes which were developed, based on related literatures. Participants were asked to illustrate three given texts using free description and query. This setting was designed to investigate a 'category search behavior' which was defined as 'the user has no specific image in mind but is able to specify requirements or conditions for the resulting image. The result will be the class of images that satisfy the conditions'. (Hollink, et al. 2004: 614). By comparing semantically richer, free descriptions and limited search queries, the authors found that terms in the general attribute were the most frequently used in both tasks; however, in the query tasks, more specific terms and less abstract and perceptual terms were found.

Rorissa (2005) investigated the relationships between common and distinctive features of images and their similarity as perceived by human judges. Common and distinctive features of images, which were derived by analysing participants' description tasks, were compared with similarity judgment and confirmed the relationships between common and distinctive features and image similarity; particularly, common features had more weight than distinctive features.

Studies have also employed a sorting task. Börner (2000) conducted a usability study comparing a latent semantic analysis algorithm applied to the information visualizer and human free-sorting task. Data analysis results showed a similarity between human sorting patterns and features derived from textual descriptions through the algorithm. Rorissa and Hastings posited that although categorization had been recognized as a way of facilitating browsing, there is no answer as to which attributes should be used to categorize images. By investigating human sorting tasks, the authors concluded that 'interpretive attributes are better candidates than perceptual attributes for indexing categories/groupings of images'. (Rorissa and Hastings 2004: 360).

Rogowitz et al. (1998) conducted two experiments to investigate how humans judge image similarity, the table scale and the computer scale and image similarity results were compared with two algorithmic image similarity metrics: a colour-histogram and a multi-resolution framework of colour, contraction and orientation-selective attributes. Results demonstrated that different approaches used in the two experiments produced similar patterns. Also, the colour of the images was significantly related to human similarity judgment. Although the authors named table scale and computer scale as similarity tasks, the setting of the table scale task has been commonly used for the sorting task. Therefore, the study results can be interpreted to mean that sorting and similarity judgment tasks show similar results.

<table>
<thead>
<tr>
<th>Study</th>
<th>Research design</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Purpose</td>
</tr>
<tr>
<td></td>
<td>Collection</td>
</tr>
<tr>
<td></td>
<td>Participants</td>
</tr>
<tr>
<td></td>
<td>Procedure</td>
</tr>
<tr>
<td></td>
<td>Purpose</td>
</tr>
<tr>
<td></td>
<td>Collection</td>
</tr>
<tr>
<td></td>
<td>Participants</td>
</tr>
<tr>
<td></td>
<td>Procedure</td>
</tr>
<tr>
<td></td>
<td>Task</td>
</tr>
<tr>
<td></td>
<td>Purpose</td>
</tr>
<tr>
<td></td>
<td>Collection</td>
</tr>
<tr>
<td>Jörgensen (1995)</td>
<td><strong>Participants</strong></td>
</tr>
<tr>
<td>------------------</td>
<td>------------------</td>
</tr>
<tr>
<td><strong>Procedure</strong></td>
<td>Descriptive viewing task: participants described six images which were projected for two minutes; Descriptive search task: participants described the six images to a librarian or an ideal image retrieval system accepting natural language; Descriptive memory task: participants described the images five weeks later; Sorting task: Participants sorted seventy-seven images into groups which would be used later for finding images; Concept search task: participants selected images which expressed two given abstract concepts, 'mysterious' and 'transformation'.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Laine-Hernandez &amp; Westman (2006)</th>
<th><strong>Task</strong></th>
<th>Describing; Sorting</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Purpose</strong></td>
<td>To evaluate current image indexing framework for reportage photographs and to find the effect of different tasks on image description and image sorting</td>
<td></td>
</tr>
<tr>
<td><strong>Collection</strong></td>
<td>Forty reportage-type photographs from two online image collections by image journalists and amateur photographers were selected based on several criteria, such as a variety of colourfulness, lightness, viewing distance and topical and emotional content.</td>
<td></td>
</tr>
<tr>
<td><strong>Participants</strong></td>
<td>Twenty native Finnish speakers (students of technology and university employees)</td>
<td></td>
</tr>
<tr>
<td><strong>Procedure</strong></td>
<td>Keywording task: participants wrote down five words for each photograph without time limitations. Free description task: participants described each photograph as they would describe its contents to another person, without time limitations. Categorization task: participants organized printed photographs.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hollink, Schreiber, Wielinga &amp; Worring (2004)</th>
<th><strong>Task</strong></th>
<th>Describing; Search</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Purpose</strong></td>
<td>To understand how people perceive images and to develop a framework for the classification of image attributes</td>
<td></td>
</tr>
<tr>
<td><strong>Collection</strong></td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td><strong>Participants</strong></td>
<td>Thirty participants (students of the University of Amsterdam and their family and friends)</td>
<td></td>
</tr>
<tr>
<td><strong>Procedure</strong></td>
<td>Participants were given three texts providing different search contexts (a paragraph from a children's book, a few lines from a historical novel and a paragraph from a newspaper) and asked to form an image in their minds illustrating each text, to write down a free text description of the images and to search for the images using a maximum of five queries.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rorissa (2005)</th>
<th><strong>Task</strong></th>
<th>Describing; Similarity judgment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Purpose</strong></td>
<td>To investigate the relationships between common and distinctive features of images and their similarity judged by human observers</td>
<td></td>
</tr>
<tr>
<td><strong>Collection</strong></td>
<td>A set of 30 colour images obtained from a book by O'Connor and Wyatt (2004)</td>
<td></td>
</tr>
<tr>
<td><strong>Participants</strong></td>
<td>150 students of the School of Library and Information Science (seventy-five participants were assigned to each of two tasks)</td>
<td></td>
</tr>
<tr>
<td><strong>Procedure</strong></td>
<td>Description task: participants describe features of each image for ninety seconds. Similarity judgment task: using magnitude estimation, participants conducted the similarity judgment task twice with 435 pairs of the thirty images (the set used for the second time was obtained by reversing the order of image pairs used for the first time).</td>
<td></td>
</tr>
<tr>
<td><strong>Task</strong></td>
<td>Sorting</td>
<td></td>
</tr>
<tr>
<td><strong>Purpose</strong></td>
<td>To conduct a usability study of the data analysis algorithm applied in the Digital Library Visualizer. Sorting task results and the Latent Semantic Analysis results from the textual image descriptions were compared.</td>
<td></td>
</tr>
</tbody>
</table>
Börner (2000) | Collection | Four image data sets obtained from the Dido image data bank (a digital library at the Indiana University Department of the History of Art) using the following four search queries, Bosch, African and two Chinese. The sets contained twelve, seventeen, thirty-one and thirty-two images, respectively.  

Participants | Twenty graduate students  

Procedure | Each participant sorted two image sets, labelled the groups and described the sorting criteria they used.  


Purpose | To investigate how natural categorization behaviour can be applied to image indexing and classification systems  

Collection | A set of fifty colour images selected from the Hemera Photo Object Volume 1.  

Participants | Thirty graduate students at a major southwestern United States university  

Procedure | Participants sorted a set of printed images without any constraints on the time or the number of categories.  

Rogowitz, Frese, Smith, Bouman, & Kalin (1998) | Task | Sorting; Similarity judgment  

Purpose | To investigate how humans judge image similarity  

Collection | Ninety-seven photographs representing various semantic categories, viewing distances and colours  

Participants | Fifteen volunteer observers from the T.J. Watson Research Center, Hawthorne Laboratory  

Procedure | Table scaling: participants arranged a set of printed images based on their similarities – similar images would be close to each other. Computer scaling: participants were presented with a reference image and eight randomly-chosen images on a computer screen and of the eight images they were asked to select the image most similar to the reference image.  

Table 1: Studies adopting image related tasks

In addition to these studies, there are studies of the extraction of emotional features through low-level features of an image (see Wang and He's (2008) survey paper). Recently, Schmidt and Stock (2009) demonstrated that scroll bars adapted to collect users' basic emotional judgments on images can be used in emotion indexing of images.

Research design

This study explored emotional perception of an image through sorting, describing and searching, which were adopted for understanding general image perceptions. In order to explore emotional perception, an emotion-space should be defined. However, in addition to a variety of emotional terms, definitions and boundaries of emotional terms are fuzzy and vague; therefore, in this study basic emotions are used in selecting images. This idea is based on basic theory, which asserted and demonstrated that there are basic level categories which include most frequently and commonly used concepts or terms (Rosch 1977). Many theorists agree there is a set of basic emotions, but there is little consensus on the number of basic emotions and what basic emotions are (Ortony and Turner 1990). The current study adopted the emotional terms and their hierarchical structures which were proposed by Shaver, et al. (1987). Shaver et al. collected candidate emotional terms (nouns) from related literatures and asked participants to rate whether each term was an emotional term or not. Then, the final selected 135 emotional words were categorized by participants in order to reveal hierarchical clusters. The hierarchy included three levels of emotions: the first level consisted of two emotional categories, positive and negative emotions which were represented by Joy and Sadness, the second level consisted of six emotional categories and the third level consisted of twenty-five emotional categories (Table 2). They found that the second level was consistent with other studies which demonstrated features of basic level emotions. The six basic emotional terms employed by Shaver et al. were used in selecting images and analysing collected data. A set of images was selected from Flickr, in which ordinary users' casual photos are uploaded, rather than using a standard dataset (such as a data set from the International Affective Picture System), which includes strong stimuli for inducing intensive affective reactions. Thirty creative commons licensed Flickr images tagged with one of the six basic emotions were selected. Then, by conducting a pilot study with eleven
graduate student participants, twelve photos (two photos for each basic emotion) were chosen. These photos were consistently given basic emotional terms in the pilot study.

<table>
<thead>
<tr>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive (joy)</td>
<td>Love</td>
<td>Affection, lust, longing</td>
</tr>
<tr>
<td>Joy</td>
<td>Cheerfulness, zest, contentment, pride, optimism, enthrallment, relief</td>
<td></td>
</tr>
<tr>
<td>(Surprise)</td>
<td>(Surprise)</td>
<td></td>
</tr>
<tr>
<td>Negative (sadness)</td>
<td>Anger</td>
<td>Irritation, exasperation, rage, disgust, envy, torment</td>
</tr>
<tr>
<td>Sadness</td>
<td>Suffering, sadness, disappointment, shame, neglect, sympathy</td>
<td></td>
</tr>
<tr>
<td>Fear</td>
<td>Horror, nervousness</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Three levels of emotional categories

Three image related tasks, sorting, describing and searching, were assigned to participants. Although this study used twelve images, for the sorting task participants were asked to sort the thirty images (4" x 6" photographs), which were used in the pilot study, into two to twenty-five groups according to emotional impressions and to give a label to each group using 135 emotional names. The sorting task used thirty images because twelve images might lead participants to generate a small number of groups. For the describing task, each of the twelve images was projected on to a large screen. Participants were asked to choose descriptors from the list of 135 emotional names for each image and then select one emotional name that described the image most appropriately. For the searching task, six searching situations were given to participants. For example, the researcher asked participants 'Let's assume that you send a query "Love" to the image search engine. Which photos are appropriate for this query?' Then, participants were asked to pick two appropriate photographs and then select the most appropriate from those two. The six basic emotions were used in the searching situations. The survey questionnaire was administered to fifty-nine students at the University of South Florida. The majority of participants were female (70.7%), undergraduate students (94.6%), between eighteen and twenty-five years old (77.6%). Their majors varied and included mass communication, world language, English, social work, criminology, biology, nursing, biomedical science and engineering.

Findings

Representation of emotion-space from the sorting and describing tasks

Of the three tasks explored in the current study, the searching task simulated a real searching situation, although it had limitations in revealing subtle and hidden information needs and their contexts, whereas the sorting and describing tasks did not directly reflect searching behaviour, although they attempted to examine an emotion-space and perceptions obtained from images. In this section, an emotion-space is examined and compared with reference to the sorting and describing tasks.

Sorting task

For the sorting task, the average number of groups generated by a participant was 6.88 and the maximum and minimum numbers of groups were 15 and 3, respectively. Approximately 40% of participants created 5 or 6 groups, and 30% of participants generated 8 or 10 groups (Table 3).
A notable finding is that although the thirty images were grouped into six sets of five images, each tagged with one of six basic emotions, the six emotions were not evenly represented through the sorting tasks. For example, as shown in Table 4, in the case that three groups were made from thirty images, those groups were not from three different basic emotions but from two basic emotions: in other words, two groups belonged to one basic emotion (two groups belonged to joy and the other group belonged to sadness). When five groups were made, these five groups were from three or four different basic emotions rather than five different basic emotional groups. Overall, when six or fewer groups were made, three or four basic emotions were mainly adopted and when more than six groups were made, five basic emotions were primarily adopted.

Table 3: Number of categories generated in sorting

<table>
<thead>
<tr>
<th>No.</th>
<th>Category 1</th>
<th>Category 2</th>
<th>Category 3</th>
<th>Category 4</th>
<th>Category 5</th>
<th>Category 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>5</td>
<td>8.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>14.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>9</td>
<td>15.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.8</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.8</td>
</tr>
<tr>
<td>Total</td>
<td>57</td>
<td>100.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Relations between the number of groups and basic emotions

<table>
<thead>
<tr>
<th>No. of groups</th>
<th>No. of basic emotions</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>1 3</td>
</tr>
<tr>
<td>5</td>
<td>4 7</td>
</tr>
<tr>
<td>6</td>
<td>1 5 3</td>
</tr>
<tr>
<td>7</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1 2 3</td>
</tr>
<tr>
<td>9</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1 1 4 1</td>
</tr>
<tr>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3 7 22 12 1</td>
</tr>
</tbody>
</table>

Table 5 presents the categories of basic emotional terms that were frequently used in the sorting task. When two categories of basic emotions appeared, regardless of the number of groups, joy and sadness were the categories adopted during the sorting task. Then, as another basic emotional category is added, one additional basic emotion was obviously noticed, such as love, joy and sadness (four out of six occurrences of three basic emotional categories), love, joy, sadness and fear or love, joy, sadness and anger (ten and five out of twenty-one occurrences of four basic emotional categories) and love, joy, sadness, anger and fear (nine out of twelve occurrences of five emotional categories). As addressed above, when groups were made, the six basic emotions were not evenly adopted; some basic emotional categories rarely appear but other categories tend to be further divided. For instance, analysing the sorting tasks that generated six groups demonstrated that most participants created one group from the love category, two groups from joy, two from sadness and one from fear.

<table>
<thead>
<tr>
<th>No. of basic emotions</th>
<th>Love</th>
<th>Joy</th>
<th>Surprise</th>
<th>Anger</th>
<th>Sadness</th>
<th>Fear</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
</tbody>
</table>
Table 5: Distribution of categories of basic emotional terms

The frequency of occurrence of emotional terms was analysed. Of 135 emotional terms from the list given by Shaver et al. (1987), 99 were used for labelling sorted categories. Of these 99 terms, 18 terms which occurred seven or more times accounted for 50% of the total occurrences of labelling terms (Table 6). Further analysis is explained below by comparing with the describing task.

<table>
<thead>
<tr>
<th>Emotional term</th>
<th>Frequency</th>
<th>%</th>
<th>Cumulative %</th>
<th>Basic category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happiness</td>
<td>20</td>
<td>5.1</td>
<td>5.1</td>
<td>Joy</td>
</tr>
<tr>
<td>Excitement</td>
<td>18</td>
<td>4.6</td>
<td>9.7</td>
<td>Joy</td>
</tr>
<tr>
<td>Sadness</td>
<td>16</td>
<td>4.1</td>
<td>13.8</td>
<td>Sadness</td>
</tr>
<tr>
<td>Loneliness</td>
<td>16</td>
<td>4.1</td>
<td>17.9</td>
<td>Sadness</td>
</tr>
<tr>
<td>Love</td>
<td>15</td>
<td>3.8</td>
<td>21.7</td>
<td>Love</td>
</tr>
<tr>
<td>Contentment</td>
<td>13</td>
<td>3.3</td>
<td>25.0</td>
<td>Joy</td>
</tr>
<tr>
<td>Thrill</td>
<td>12</td>
<td>3.1</td>
<td>28.1</td>
<td>Joy</td>
</tr>
<tr>
<td>Tenseness</td>
<td>10</td>
<td>2.6</td>
<td>30.7</td>
<td>Fear</td>
</tr>
<tr>
<td>Affection</td>
<td>9</td>
<td>2.3</td>
<td>33.0</td>
<td>Love</td>
</tr>
<tr>
<td>Enjoyment</td>
<td>8</td>
<td>2.0</td>
<td>35.0</td>
<td>Joy</td>
</tr>
<tr>
<td>Exhilaration</td>
<td>8</td>
<td>2.0</td>
<td>37.0</td>
<td>Joy</td>
</tr>
<tr>
<td>Isolation</td>
<td>8</td>
<td>2.0</td>
<td>39.0</td>
<td>Sadness</td>
</tr>
<tr>
<td>Fear</td>
<td>8</td>
<td>2.0</td>
<td>41.0</td>
<td>Fear</td>
</tr>
<tr>
<td>Caring</td>
<td>7</td>
<td>1.8</td>
<td>42.8</td>
<td>Love</td>
</tr>
<tr>
<td>Amusement</td>
<td>7</td>
<td>1.8</td>
<td>44.6</td>
<td>Joy</td>
</tr>
<tr>
<td>Bliss</td>
<td>7</td>
<td>1.8</td>
<td>46.4</td>
<td>Joy</td>
</tr>
<tr>
<td>Amazement</td>
<td>7</td>
<td>1.8</td>
<td>48.2</td>
<td>Surprise</td>
</tr>
<tr>
<td>Uneasiness</td>
<td>7</td>
<td>1.8</td>
<td>50.0</td>
<td>Fear</td>
</tr>
</tbody>
</table>

Table 6: Frequency of emotional terms used in the sorting task

<table>
<thead>
<tr>
<th>Emotional term</th>
<th>Frequency</th>
<th>%</th>
<th>Cumulative %</th>
<th>Basic category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happiness</td>
<td>20</td>
<td>5.1</td>
<td>5.1</td>
<td>Joy</td>
</tr>
<tr>
<td>Excitement</td>
<td>18</td>
<td>4.6</td>
<td>9.7</td>
<td>Joy</td>
</tr>
<tr>
<td>Sadness</td>
<td>16</td>
<td>4.1</td>
<td>13.8</td>
<td>Sadness</td>
</tr>
<tr>
<td>Loneliness</td>
<td>16</td>
<td>4.1</td>
<td>17.9</td>
<td>Sadness</td>
</tr>
<tr>
<td>Love</td>
<td>15</td>
<td>3.8</td>
<td>21.7</td>
<td>Love</td>
</tr>
<tr>
<td>Contentment</td>
<td>13</td>
<td>3.3</td>
<td>25.0</td>
<td>Joy</td>
</tr>
<tr>
<td>Thrill</td>
<td>12</td>
<td>3.1</td>
<td>28.1</td>
<td>Joy</td>
</tr>
<tr>
<td>Tenseness</td>
<td>10</td>
<td>2.6</td>
<td>30.7</td>
<td>Fear</td>
</tr>
<tr>
<td>Affection</td>
<td>9</td>
<td>2.3</td>
<td>33.0</td>
<td>Love</td>
</tr>
<tr>
<td>Enjoyment</td>
<td>8</td>
<td>2.0</td>
<td>35.0</td>
<td>Joy</td>
</tr>
<tr>
<td>Exhilaration</td>
<td>8</td>
<td>2.0</td>
<td>37.0</td>
<td>Joy</td>
</tr>
<tr>
<td>Isolation</td>
<td>8</td>
<td>2.0</td>
<td>39.0</td>
<td>Sadness</td>
</tr>
<tr>
<td>Fear</td>
<td>8</td>
<td>2.0</td>
<td>41.0</td>
<td>Fear</td>
</tr>
<tr>
<td>Caring</td>
<td>7</td>
<td>1.8</td>
<td>42.8</td>
<td>Love</td>
</tr>
<tr>
<td>Amusement</td>
<td>7</td>
<td>1.8</td>
<td>44.6</td>
<td>Joy</td>
</tr>
<tr>
<td>Bliss</td>
<td>7</td>
<td>1.8</td>
<td>46.4</td>
<td>Joy</td>
</tr>
<tr>
<td>Amazement</td>
<td>7</td>
<td>1.8</td>
<td>48.2</td>
<td>Surprise</td>
</tr>
<tr>
<td>Uneasiness</td>
<td>7</td>
<td>1.8</td>
<td>50.0</td>
<td>Fear</td>
</tr>
</tbody>
</table>
For the describing task, participants were asked to provide as many emotional terms from the list as they wanted to (task D1) and to select the most appropriate emotional term (task D2). When participants were allowed to assign as many terms as they wanted, on average they provided 3.52 terms per image and the maximum and minimum number of terms were 9 and 1, respectively. Although it is beyond the scope of the research objectives, factors related to research design were examined. First, to determine whether the order of descriptors in D1 is related to selecting the most appropriate emotional term. From 633 describing task results, 31.12% (197 describing tasks) selected the first descriptors for the D2 task, 32.39% (205 describing tasks) used any of descriptors in the middle and 26.07% (165 describing tasks) used the last descriptor. 4.90% of participants (31 describing tasks) provided only one descriptor during the D1 task and 5.53% (35 describing tasks) did not follow the instructions given and provided a new term in D2. This result shows that the order of descriptors given by users is not related to the significance of descriptors.

Secondly, to determine whether the order of describing tasks affects the number of descriptors. As shown in Figure 1, except for the last image, which has remarkably few descriptors, the order of projected images seems to have no relation to the number of descriptors.

<table>
<thead>
<tr>
<th>Task D1</th>
<th>Task D2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Emotional term</strong></td>
<td><strong>Frequency</strong></td>
</tr>
<tr>
<td>Joy</td>
<td>76</td>
</tr>
<tr>
<td>Happiness</td>
<td>73</td>
</tr>
<tr>
<td>Enjoyment</td>
<td>68</td>
</tr>
<tr>
<td>Amusement</td>
<td>59</td>
</tr>
<tr>
<td>Love</td>
<td>55</td>
</tr>
<tr>
<td>Delight</td>
<td>52</td>
</tr>
<tr>
<td>Caring</td>
<td>51</td>
</tr>
<tr>
<td>Affection</td>
<td>50</td>
</tr>
</tbody>
</table>
Table 7: Frequency of emotional terms used in the describing task

Distributions of basic emotional categories which appeared in the sorting and two describing tasks were compared. As shown in Table 8, overall distributions were similar over three tasks; emotional terms which belong to the joy category were most dominant followed by terms in the sadness, love, fear, anger and surprise categories.

Table 8: Categorical distribution of emotional terms used in the sorting and describing tasks

The participants were instructed to select the terms from the given list of 135 emotional terms, but some participants used their own terms in the sorting and describing tasks. They said that they could not find appropriate terms from the list and they had a better term in mind. Terms that appeared twice or more were: relaxed (relaxing), beautiful, cold, fun, calm, wet, tired, scared, dull, boring, indifference and serenity. Since this study was designed to use 135 emotional terms, it was not the scope of this study to investigate these unlisted terms. However, they would demonstrate users' understanding of emotional scope, so further discussion is provided in the next section.
Comparison of individual image perceptions across three tasks

This section examines whether the three tasks (sorting, describing, and searching) represent emotional perception of an image differently. As presented in Table 9, of twelve images, six (L1, J1, J2, A2, Sa2 and F2) showed consistent distributions across three tasks and four (L2, Su1, Su2 and Sa1) were perceived differently in the searching task than the other two tasks. Two images (A1 and F1) were perceived differently in each of the three tasks. The first notable finding is that half of the twelve images demonstrated different distribution between the searching and sorting or describing tasks. A possible explanation is that the searching task adopted in this study forced participants to make relative judgments among images, whereas the other two tasks asked them to describe and sort images freely. For instance, although an image (L2) conveys an impression of love as well as joy, if the image is perceived as the most joyful among the other images, the image has a high rating for joy in the searching task. However, the same image can be sorted and described mainly as an image expressing love, because participants perceived that the image itself expressed love rather than joy. Also, the images (Su1 and Su2) which were tagged with surprise were rarely described or sorted as a surprise emotion, but when participants were asked to select images expressing surprise, these two images were chosen. Furthermore, ten out of twelve images demonstrated similar distributions between the sorting and describing tasks, in general. However, a closer observation revealed that task D2 tends to present a more dominant emotional perception of an image than the D1 or sorting tasks and the sorting task presented more diverse emotional perceptions related to an image.

<table>
<thead>
<tr>
<th></th>
<th>L1</th>
<th>L2</th>
<th>J1</th>
<th>J2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>S</td>
<td>D1</td>
<td>D2</td>
</tr>
<tr>
<td>Love</td>
<td>100.0</td>
<td>57.9</td>
<td>74.7</td>
<td>86.8</td>
</tr>
<tr>
<td>Joy</td>
<td>0.0</td>
<td>33.4</td>
<td>20.7</td>
<td>13.2</td>
</tr>
<tr>
<td>Surprise</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Anger</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Sadness</td>
<td>0.0</td>
<td>3.6</td>
<td>0.5</td>
<td>0.0</td>
</tr>
<tr>
<td>Fear</td>
<td>0.0</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Su1</th>
<th>Su2</th>
<th>A1</th>
<th>A2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>S</td>
<td>D1</td>
<td>D2</td>
</tr>
<tr>
<td>Love</td>
<td>0.0</td>
<td>5.3</td>
<td>2.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Joy</td>
<td>5.0</td>
<td>70.2</td>
<td>57.9</td>
<td>50.9</td>
</tr>
<tr>
<td>Surprise</td>
<td>95.0</td>
<td>7.1</td>
<td>13.8</td>
<td>28.3</td>
</tr>
<tr>
<td>Anger</td>
<td>0.0</td>
<td>0.0</td>
<td>5.1</td>
<td>3.8</td>
</tr>
<tr>
<td>Sadness</td>
<td>0.0</td>
<td>1.8</td>
<td>2.0</td>
<td>3.8</td>
</tr>
<tr>
<td>Fear</td>
<td>0.0</td>
<td>10.5</td>
<td>13.3</td>
<td>11.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Sa1</th>
<th>Sa2</th>
<th>F1</th>
<th>F2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>S</td>
<td>D1</td>
<td>D2</td>
</tr>
<tr>
<td>Love</td>
<td>7.1</td>
<td>0.0</td>
<td>3.3</td>
<td>3.8</td>
</tr>
<tr>
<td>Joy</td>
<td>0.0</td>
<td>1.8</td>
<td>2.5</td>
<td>1.9</td>
</tr>
<tr>
<td>Surprise</td>
<td>2.4</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Anger</td>
<td>85.7</td>
<td>3.6</td>
<td>3.4</td>
<td>3.8</td>
</tr>
<tr>
<td>Sadness</td>
<td>0.0</td>
<td>87.9</td>
<td>83.6</td>
<td>86.8</td>
</tr>
<tr>
<td>Fear</td>
<td>4.8</td>
<td>5.3</td>
<td>4.4</td>
<td>1.9</td>
</tr>
</tbody>
</table>

* Image IDs are based on the tag on the images; L (Love), J (Joy), Su (Surprise), A (Anger), Sa (Sadness), F (Fear)
The tasks are identified as: F for searching (finding the one most appropriate image for each
Table 9: Overall differences across three tasks

Discussion

Based on previous studies which demonstrated that emotional perceptions of images are well presented through image related tasks rather than search queries (Jörgensen 1996; O'Connor et al. 1999; Greisdorf and O'Connor 2002; Laine-Hernandez and Westman 2006), the current study examined how emotional perceptions of images are represented through sorting, describing and searching tasks. First, it was found that joy and sadness were dominantly perceived through the sorting and describing tasks. During the sorting task, image viewers tended to distinguish images based on whether images are joyful or sad and then subdivide joyful or sad emotions rather than adopt other basic emotions. In contrast to joy and sadness, surprise was rarely recognized during these two tasks. These results are in accordance with the study by Shaver et al. (1987) which identified two fundamental emotions, positive and negative (or joyful and sad) and had reservations concerning the surprise emotion. Wild, Erb and Bartels (2001) also demonstrated that happiness, which belongs to the joy category in the study by Shaver et al. and sadness are significantly evoked by facial expression. This finding proposes a potential that an emotion-space, which other disciplines (such as psychology, cognitive science and so on) have explored within general environments or using other stimuli, may be applied to an image retrieval environment. Therefore, it would be worthwhile effort to revisit and apply those theories on emotions, with the purpose of improving image retrieval effectiveness.

Although the current study provided participants with a list of comprehensive emotional terms, 20.75% of describing task participants and 17.54% of sorting task participants used terms not listed in the given list. As Ortony and Turner (1990) addressed, there is no entire agreement on the definition and scope of emotions. Therefore, some emotions which are included in one researcher's list are not included in others. For example, Averill (1975) included 558 words in his list, but Shaver et al. (1987) included only 135 terms, even though they considered Averill's list. Of the terms not included by Shaver et al. but which were selected twice or more by participants (relaxed (relaxing), beautiful, cold, fun, calm, wet, tired, scared, dull, boring, indifference and serenity), it is debatable whether some are emotions (for example, boring and serenity are in Davitz's (1969) list but not in Shaver's list). While terms describing cognitive status are arguable, terms describing objects' status or atmosphere (beautiful, cold and wet) are obviously non-emotional terms.

However, even though theoretically the terms describing objects' status or atmosphere are not included in the list of emotions, if they can be used for providing emotional access they should be considered in image indexing and retrieval systems. Another finding which should be considered in image indexing is that 15% of emotional terms (approximately 20 out of 135) accounted for 50% of the total occurrences of descriptions and labels during the describing and sorting tasks. If it is practically impossible to index a number of subjective and versatile emotional terms during the representation process, terms which are frequently and popularly adopted by users should receive priority. Once images are indexed with popular emotional (and emotion-like) terms, browsing functions which guide users to navigate an emotion-space would help users find images satisfying more specialized and detailed emotional needs.

Further research is needed to decide which emotional terms can be entry points and how those entry points can be expanded or explored while searching. For example, a controlled vocabulary system which includes keyemotional terms and their related emotional terms can be developed and evaluated in terms of image retrieval effectiveness.

Finally, characteristics of the sorting, describing and searching tasks were analysed based on differences of individual image perceptions. Whereas the sorting and describing tasks demonstrated relatively high accordance in emotional perception of an image, the searching task showed different patterns. The author explained the result by the research setting: whereas the sorting and describing tasks asked participants to freely describe and sort images, the searching task forced participants to make relative judgments among images. Since the searching task depends on a relatively small number of data sets, this task may not represent major emotional perceptions of an individual image precisely, even when a participant perceived a strong joyful impression from an image, if the image was the most appropriate for love from twelve images, the participant had to select the image for the searching task looking for love.
However, if a searching task employs an image retrieval system with a reasonably large number of images, the author assumes that the degree of relativity may decrease, which may result in more precise emotional perceptions of images. Based on this finding, it is suggested that when developing a research design it should be taken into account that perceptions of an image can be different depending on the research constraints and whether the image is viewed by itself or compared with other images.

Another finding is that the sorting task demonstrated more diverse emotional perceptions of one image. Jörgensen (1996) explained the difference between the describing task, in which participants focus on individual images and the sorting task, in which participants see a whole set of images. This could be one explanation for that finding. Another reason that may be inferred is that more personal background, preference and intellectual processing are included during the sorting task than in the describing task and this might result in a more diverse interpretation of an image. Therefore, the sorting task seems more appropriate than other tasks to explore the multi-faceted aspects of emotional meanings.

Conclusion

This study explored emotional perceptions of image documents through three image related tasks: sorting, describing and searching. The sorting and describing tasks demonstrated that joy and sadness are two major categories of emotions which people perceive from images. It was also found there is only a relatively small number of emotional terms frequently chosen as descriptors or labels. It was suggested, therefore, that by assigning emotional index terms using a small number of popular emotional terms, image retrieval effectiveness can be improved with minimum indexing efforts. When comparing image perceptions of an individual image across three tasks, overall similar perceptions were discovered across the sorting, describing and searching tasks.

However, it was noticed that minor differences are caused by the unique characteristics of each task. Therefore, features of each task should be considered in any research design which explores image perceptions. Since a limitation of this study is that image related tasks were conducted with the small number of images by undergraduate students, further research is needed to test these results with a large data set and a more diverse population. Differences of emotional perceptions in terms of demographic features, such as sex, age and ethnicity, might be a topic of future research. In addition, although the current study focused on affective perception of an image, a similar approach could be applied to overall semantic perceptions of images.

Acknowledgements

This work was supported, in part, by University of South Florida Internal Awards Program under Grant No. R061581. The author thanks to Ms. Jyoti Deo, a master's student at the University of South Florida, for her assistance in data collection and analysis. The author also thanks to copy-editors of the journal and Ms. Vickie Toranzo Zacker, a master's student at the University of South Florida, for their assistance in enabling the author to satisfy the style requirements of the journal.

About the author

JungWon Yoon is an Assistant Professor in the School of Information, University of South Florida. She received her Bachelor's degree in Library and Information Science and Master of Library and Information Science from Ewha Womans University, Seoul, Korea and her PhD from the University of North Texas. She can be contacted at: jyoon@usf.edu.

References


Appendix