

2010

# A New Model For Semantic Photograph Description Combining Basic Levels and User-assigned Descriptors

Hyuk-Jin Lee

Texas Woman's University, [hlee@twu.edu](mailto:hlee@twu.edu)

Diane M. Neal

The University of Western Ontario, [dneal2@uwo.ca](mailto:dneal2@uwo.ca)

Follow this and additional works at: <https://ir.lib.uwo.ca/fimspub>

 Part of the [Library and Information Science Commons](#)

---

## Citation of this paper:

Lee, Hyuk-Jin and Neal, Diane M., "A New Model For Semantic Photograph Description Combining Basic Levels and User-assigned Descriptors" (2010). *FIMS Publications*. 18.

<https://ir.lib.uwo.ca/fimspub/18>



# A new model for semantic photograph description combining basic levels and user-assigned descriptors

**Hyuk-Jin Lee**

*School of Library and Information Studies, Texas Woman's University, Texas, U.S.A.*

**Diane Neal**

*Faculty of Information and Media Studies, The University of Western Ontario, Ontario, Canada*

## Abstract

Few studies have been conducted to identify users' desired semantic levels of image access when describing, searching, and retrieving photographs online. The basic level, or the level of abstraction most commonly used to describe an item, is a cognitive theory currently under consideration in image retrieval research. This study investigates potential basic levels of description for online photographs by testing the Hierarchy for Online Photograph Representation (HOPR) model, which is based on a need for a model that addresses users' basic levels of photograph description and retrieval. We developed the HOPR model using the following three elements as guides: the most popular tags of all time on Flickr, the Pyramid model for visual content description by Jørgensen, Jaimes, Benitez, and Chang, and the nine classes of image content put forth by Burford, Briggs, and Eakins. In an exploratory test of the HOPR model, participants were asked to describe their first reaction to, and possible free-text indexing terms for, a small set of personal photographs. Content analysis of the data indicated a clear set of user preferences that are consistent with prior image description studies. Generally speaking, objects in the photograph, and events taking place in the photograph, were the most commonly used levels of description. The preliminary HOPR model shows promise for its intended utility, but further refinement is needed through additional research.

Keywords: Image Retrieval; Image Indexing; Entry Point

## 1. Introduction and related research

The need for digital images to be represented, searched, and retrieved satisfactorily online is apparent. Millions of images are uploaded to Flickr and other photograph sharing Web sites every day. People post photographs on their social networking profiles to share them with friends and family. Commercial and non-profit Web sites provide useful images to their users. In the last two decades, designers and researchers have made an

## Hyuk-Jin Lee and Diane Neal

extraordinary amount of progress in improving access within a short time span that has coincided with significant technological development [1]. However, satisfactory search results continue to elude the online community. Issues such as a lack of consistent image metadata, low precision on Web-based image search engines, and gaps in understanding of how users search for images on the Web make it difficult for people to find their desired visual content.

Many studies have demonstrated the difficulty of describing what an image is ‘about,’ especially for more abstract levels of description. Panofsky’s [2] often-cited model outlined three levels of abstraction for describing art content: pre-iconographical, iconographical, and iconological. As Shatford [3] noted, Panofsky’s iconographical and iconological levels cannot be indexed with any degree of consistency due to their subjective nature as well as openness to individual interpretation. She emphasized that indexing Aboutness should not be dismissed because of its subjective nature, however, and in indexing a general collection, each picture should be indexed for Ofness or Aboutness as appropriate. This is because the semantic analysis of an image typically identifies multiple layers of meaning [4]. Burford, Briggs, and Eakins [5] believe that Panofsky’s model is too entrenched in art history to be generalizable to image retrieval as a whole, and that the three categories are too broad to apply to the spectrum of potential contexts as necessary.

Adding to this criticism, librarians or other professionals may tend to only describe the Ofness of an image when indexing it [6], although users may describe images using terms that do not describe items apparent in the image [7]. Image query analyses have uncovered further problems. For example, many unique terms exist in Web image queries [8], and there is a wide range of query types not documented in previous predominant image query classification models [9]. This far-reaching conundrum of subjectivity and variety, both in image description and in image search, contributes to the challenge of the image retrieval task.

The vast collection of images on the Web makes it impossible for professionals to describe all the images available online. Additionally, pre-defined, authoritative indexing terms might be of questionable value for meeting general searchers’ needs [10-12]. Besser [13] pointed out that text-based image retrieval systems have not met user needs because they do not provide the multitude of access points that an individual might use to retrieve an image. It could be said that a contributing factor to this issue is that pictures are not words, and that semantic meaning gets lost in the necessary translation between picture and text [12, 14]. That said, attempts at semantically-based visual search tools for visual materials have been under-developed at best (see Rorvig, Turner, & Moncada, 1999, for an example [15]).

Browsing functionality is an important, and sometimes the most expeditious, method of document access [16]. This is especially true in the case of image access, because the human eye can scan a large set of pictures very quickly to determine content [17]. However, the ability to perform keyword searches on a collection of images is also necessary. Many browsing activities begin with a keyword search [18, 19]. Other studies have found keyword searching to be the dominant access method, such as Matusiak [20]. Limitations in the indexing of institutional image collections have included a lack of exhaustivity and specificity as well as interindexer consistency [21, 22]. Bates [23] explained that the user’s experience is phenomenologically different from the indexer’s experience, an idea that further supports the need for new approaches.

Increasingly, perhaps in response to these concerns as well as to the recognition of current online user behaviours, research is backing the promise of newer methods, such as the utility of user-assigned descriptors (UADs) for image indexing and retrieval [24, 25]. It has been suggested in various forums, however, that a combination of controlled vocabulary terms and UADs might be ideal for addressing problems with each approach: the inflexibility of authoritative descriptive methods, and the inconsistency of UADs [20, 25, 26]. Following another route, automatic semantic indexing using content-based image retrieval (CBIR) methods is also showing promise in research [27, 28]. Content-based image retrieval is based on the use of low-level perceptual features present in an image such as shapes and colors [29]. However, little empirical evidence supporting user satisfaction with UADs, automatic semantic indexing, or CBIR exists, and that support alone does not provide any information regarding the most commonly used levels of semantic description.

To date, relatively few studies have been conducted to identify the “entry point” level [30] or “basic level” [31] used by image searchers. According to the notion of an “entry point,” any given object has one particular level of semantic description at which it is first processed in one’s mind. It may vary according to users’ domain

## Hyuk-Jin Lee and Diane Neal

knowledge of the items or concepts represented in a picture, but at the same time, generalizations may be needed for widespread Web use [8]. For the purposes of this study, we assume that users' first reaction to an image can be used, and is frequently used, as an entry point during the retrieval process. Similarly, the basic level of an object is the level of abstraction that is most commonly used to describe it. For example, one might identify a computer as a computer (its basic level), rather than a piece of electronic equipment (a superordinate term to computer) or a laptop (a subordinate term to computer).

The basic level theory has potential for improving general Web image searching. Unlike expert-oriented image databases, today's Web images are typically described and queried using non-controlled vocabulary terms. Often, users assign small amounts of descriptive metadata to the image based on their readily available knowledge of, or reaction to, its semantic content [32, 25]. Due in part to its thoroughly-tested cognitive validity, the basic level theory has become prominent in recent image retrieval research. Studies have indicated the tendency for people to describe individual images at the basic level of the image's content, while image collections are described at the superordinate level [33, 34]. [33] concluded after performing experiments involving sorting and labelling images as well as groups of images that "the number of subordinate, basic, and superordinate level terms have statistically significant differences" (p. 1749). Stvilia and Jørgensen [35] found that descriptions of Flickr photosets are based on shared attributes of the photographs within the photoset.

A more thorough understanding of users' preferred entry points, and their incorporation into online image retrieval systems, could replace or enhance the use of professionally-assigned authoritative indexing terms as well as UADs [25]. This is considered a necessary area of investigation due to problems of tagging inconsistency identified in previous research [19, 25]. This exploratory study aimed to identify those entry points. In the process, we have developed a proposed theoretical model, the Hierarchy for Online Photograph Representation (HOPR) model. The HOPR model was developed with online personal photograph classification in mind. While existing models of concept-based image indexing and description certainly bear worth, it seems that a model providing expansion within the basic levels of personal photograph description would be a worthwhile contribution to further refine existing models. Therefore, possible basic levels of preferred user description for personal photographs can be found within its hierarchical range of abstraction levels. A test of this model using participants' responses to personal photographs, which is described in this paper, reveals the point of abstraction and description at which the basic level of personal photograph description might lie. Future instantiations of tag-based information retrieval systems might better serve users by incorporating pointers to basic level tags for more consistent description levels.

## 2. Theoretical background

We chose two models as a basis for this study and the associated model: the Pyramid model for visual content description [36, 37], and the nine classes of image content put forth by Burford, Briggs, and Eakins [5]. These models were chosen based on their shared strengths. They both emphasize the description of image content from a non-expert point of view, and they both hold promise for practical application in image retrieval systems.

The Pyramid model was originally introduced in Jaimes and Chang [36] and further described and tested in Jørgensen, Jaimes, Benitez, and Chang [37]. Discussing prevalent problems in image indexing, Jaimes and Chang noted the myriad of visual information present in one image, as well as the many possible varying levels of description. Using an interdisciplinary approach, they sought to develop an intuitive image indexing model that distinguishes between visual and non-visual information. The ten-level hierarchical structure is based on four levels of syntactical description and six levels of semantic description. The syntactical elements are the lowest levels, and they focus on lower-level features such as production techniques, colors, textures, and so on. The semantic elements obviously focus on higher levels of meaning, from items easily identifiable present in the image at their lowest level to descriptions requiring interpretative thought. According to the authors, the Generic elements focus on the basic level categories, and the Specific elements focus on named entities. However, in reality, categorization of non abstract semantic content to generic or specific classes is a considerable oversimplification [38]. Enser [38] pointed out, "In the natural order of physical existence, an entity can be interpreted

Hyuk-Jin Lee and Diane Neal

into a hierarchy or related superconcepts and sub-concepts. As one navigates a path through such a concept hierarchy, it may not be obvious at what level one crosses the threshold from generic to specific” (p. 10). The Pyramid model appears in Figure 1.

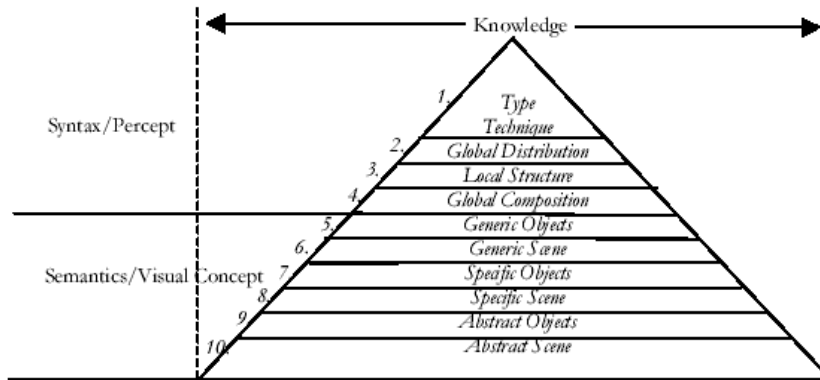


Figure 1. Pyramid Model

The second model that informed this study is Burford, Briggs, and Eakins’ nine classes of image content [5]. The authors drew from a large body of interdisciplinary sources when they created the classes, such as psychology, art, and history. Table 1 demonstrates the summary of the categories.

Table 1. Nine classes of image content, from Burford, Briggs, and Eakins (2003)

Category	Definition
Perceptual primitives	The content created by low-perceptual systems. In practical term, color and some textual descriptions which do not rely on a higher level may be categorized here.
Geometric primitives	Simple two- and three- dimensional non-representational forms, such as lines, arc, square, circle, etc.
Visual extension	Visual meaning which requires some inference, such as depth, from shadow, occlusion, perspective, etc.
Semantic units	Names, both general and specific. Most descriptions will have some naming content, though it may be subsumed in higher levels.
Contextual abstraction	Associations or interpretations which depend on environmental knowledge.
Cultural abstraction	Associations which reply on specific cultural knowledge. This may be the viewer’s own culture, or simply one of which they are aware.
Technical abstraction	Associations which reply on detailed special knowledge and vocabulary. Again this may be

## Hyuk-Jin Lee and Diane Neal

	through direct experience of an area, or second-hand knowledge.
Emotional abstraction	Emotional and affective associations. These may be generalizable, but will be filtered by the viewers' own experiences.
Metadata	Information which describes the image, but is not actual image content, such as format, size, aspect ratio, etc.

Neal [19, 25] used Burford, Briggs, & Eakins' model as a basis for her research on photojournalism professionals' online photo archive searching preferences, refining and renaming some categories, and adding categories for UADs and browsing. According to an online survey and in-person interviews, she found that the most highly preferred methods for searching were: (1) events happening in the photo, (2) objects represented in the photo, (3) UADs, (4) certain metadata about the image itself (such as the date on which the picture was shot), and (5) browsing.

### 3. Research questions

RQ1. What basic levels of Web photograph description can be ascertained from general Web users' descriptions as applied to the HOPR model (described below and printed in Appendix B)?

RQ2. What is the comparison between the HOPR model and the Pyramid model; and between the HOPR model and Burford, Briggs, and Eakins' nine classes of image content?

RQ3. How much utility is present in the HOPR model's application to online photograph searching for a generalized set of Web users?

### 4. Methodology

In order to obtain a fairly accurate but broad picture of how people are truly describing their photographs online, we independently classified the "All time most popular tags" on flickr.com (<http://www.flickr.com/photos/tags>) into superordinate categories. We then combined them into one set of eight categories: Color, Living Thing, Time, Events, Places, Concepts, Environment, and Objects. The independently created categories by the authors were similar in nature, scope, and number. We tried to approach the study from the non-image expert point of view as the major portion of the user group of Web images such as Flickr is the non-expert. Both authors are non-image experts and the subjects of this study are same as well. We chose 16 photographs (two photographs per category) from one author's personal photograph collection that seemed to best represent the categories. In a sense, we considered this selection method to be a step in understanding the subjective reaction to photographs – would our representational choices somehow correspond with the participants' descriptions? The photos appear in Appendix A.

After choosing the photos, the survey was developed and made available online using <http://www.surveymonkey.com>. For each of the 16 photos, two questions were asked of the participants. The first question was, "Provide your first, instant reaction to the content of each picture. Only spend a few seconds looking at the picture. Describe what you see in 10 words or less. Unfortunately, the SurveyMonkey tool did not provide a method of controlling how long the respondent looked at each picture; in future research, we plan to apply a different tool that can log such data. The purpose of this question was to elicit participants' entry points for the pictures. The second question, meant to gather indexing terms from the participants, was "How would you

Hyuk-Jin Lee and Diane Neal

index this picture for later retrieval by World Wide Web users? Provide your own indexing term(s).” Essentially, we elicited participants’ first responses to the content of the images (presumably, the “basic level”), and then asked them to provide free-text indexing terms, or UADs, for the pictures. The participants were not familiar with the content of the pictures, with the possible exception of photographs that display landmarks which exist in the geographical region in which many of the participants live, such as Places 1 and Living Thing 2.

To help prevent the possibility of participants determining patterns in the presented photographs, we randomized the order of the photograph display. The first eight pictures were based on the alphabetical order of the categories: Color, Concepts, Environment, Events, Living Thing, Objects, Places, and Time. The second eight pictures were presented in this order: Color, Environment, Living thing, Places, Concepts, Events, Objects, and Time. A sample survey question appears in Figure 2.

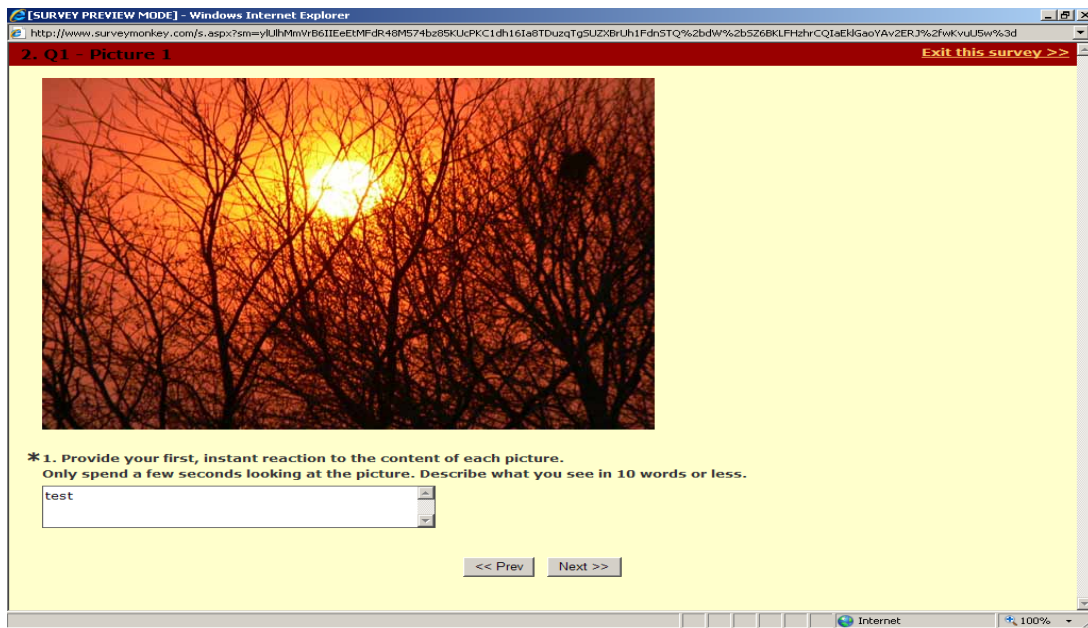


Figure 2. Sample Survey Question

Survey participants were recruited via an email invitation, and 138 library science graduate students completed the survey. Our preferred target population was “general Web users,” and the subjects of this study may represent this population as they are non-image experts and their age ranges from teens to fifties. However, we expected to be able to learn about this one sample, with a plan to expand the sample to a less homogenous population in future studies. After collecting the data, we analyzed the responses inductively, with the use of the Pyramid model as a guide, which tested the HOPR model (see Appendix B). The model, which we developed using the most popular tags of all time on Flickr, the Pyramid model for visual content description by [36, 37], and the nine classes of image content put forth by Burford, Briggs, and Eakins [6] as guides, has seven main categories and 28 subcategories. Each category and subcategory has a simple definition and examples. Several main categories are certainly guided by the Pyramid model, but the HOPR model expands on existing levels, especially in areas of the hierarchy that appear to be used frequently based on the results of this study, and adds new ones. The examples were drawn from the survey data.

To test the model, we performed content analysis. Using the HOPR model as their guide, two coders not associated with the study and unfamiliar with the images classified the data into one category or subcategory. Each coder provided the Roman numeral and the letter of the category and subcategory to each response by inputting them into a spreadsheet. For example, if the response was "Brrr! This water's freezing!" they might have assigned it the code "I.D.", representing “Abstract scene/Talk bubbles.” The coders were instructed to assign the first category in the model that they thought described the participant’s response, going down the hierarchy from

## Hyuk-Jin Lee and Diane Neal

most abstract to least abstract. Thus, the coders could assign the most relevant subcategory that best describes the response. In addition, this approach was used to identify the level of each response in the most time-efficient manner possible.

### 5. Results and discussion

Due to the large number of potential codes, reliability was a concern prior to the analysis. Cohen's kappa coefficient is a statistical measure of inter-rater agreement for qualitative (categorical) items. It is generally thought to be a more robust measure than simple percent agreement calculation since it takes into account the agreement occurring by chance. If subcategories are not included and only the seven main categories are compared, Cohen's kappa was 0.56 for Q1 ("Provide your first, instant reaction to the content of each picture. Only spend a few seconds looking at the picture. Describe what you see in 10 words or less") and 0.64 for Q2 ("How would you index this picture for later retrieval by World Wide Web users? Provide your own indexing term(s).") This indicates a moderate level of agreement for Q1, and substantial agreement for Q2 [39]. Only the codes for which both coders agreed were used for analysis.

#### 5.1. Popular basic levels and index term

To determine possible basic levels, we analyzed the most frequently selected subcategories; 150 was the lower frequency threshold for inclusion in the analysis (see Table 2). 70.9% of the Q1 codes belonged to the 10 most frequently chosen subcategories, and 74.6% of the Q2 codes belonged to the eight most frequently chosen subcategories in Q2.

Table 2. Most frequently coded categories for Q1 and Q2 all 28 subcategories

Q1: 10 Most frequent categories among 28	Q2: 8 Most frequent categories among 28
1. V.F. Generic Scene (Event) 345 (11.2%)	1. VI.D. Generic object (Artificial object) 414 (13.9%)
2. I.A. Abstract Scene (Emotion) 302 (9.8%)	2. V.F. Generic scene (Event) 380 (12.7%)
3. VI. D. Generic Object (Artificial Object) 288 (9.4%)	3. VI.E. Generic object (Human with location information) 283 (9.5%)
4. V.D. Generic Scene (Location with specific information) 213 (6.9%)	4. VI.B. Generic object (Artificial object with location information) 282 (9.5%)
5. V.C. Generic Scene (Nature Scene) 210 (6.8%)	5. V.E. Generic scene (Location) 253 (8.5%)
6. VI.E. Generic Object (Human with location information) 175 (5.7%)	6. V.C. Generic scene (Nature scene) 246 (8.2%)
7. I.B. Abstract Scene (Opinion) 173 (5.6%)	7. V.D. Generic scene (Location with specific information) 202 (6.8%)
8. VI.B. Generic Object (Artificial object with color description) 169 (5.5%)	8. VI.A. Generic object (Nature object) 165 (5.5%)
9. I.E. Abstract Scene (Talk bubbles) 160 (5.2%)	
10. V.E. Generic Scene (Location) 150 (4.9%)	

For both Q1 (basic level or entry point) and Q2 (index term), it is easy to see which HOPR levels are used most often. The results of Q1 show that VI (Generic Object, 34.64%) and V (Generic Scene, 34.12%) are the two most frequently selected categories, followed by I (Abstract Scene, 27.60%). These three categories represent 96.36% of the coded information. The other four main Q1 categories' results are as follows: II (Abstract Object, 0%), III (Specific Scene, 1.72%), IV (Specific Object, 0.10%), and VII (Physical Content, 1.82%).



## Hyuk-Jin Lee and Diane Neal

The results for Q2's seven main categories varied somewhat from Q1's. Two main categories, VI (Generic Object, 52.72%), and V (Generic Scene, 39.74%), account for 92.46% of the analyzed coded results. It should be noted that the participants did not translate their Q1 responses for I (Abstract Scene) into indexing terms. The other five main Q2 categories' results are as follows: I (Abstract Scene, 6.44%), II (Abstract Object, 0%), III (Specific Scene, 0.80%), IV (Specific Object, 0.13%), and VII (Physical Content, 0.17%).

Analyzing the three most popular main categories in Q1 (basic level) in detail by subcategory provides some interesting insights. The most popular subcategories among the three categories are: I.A. (Abstract Scene - Emotion, 9.81%), V.F. (Generic Scene - Event, 11.20%), and VI.D. (Generic Object - Artificial object, 9.35%). The participants actively described their emotions and focused on what they thought was happening in the image (events) when they saw the images in the first few seconds (basic levels). The next most frequently chosen subcategory groups were as follows: I.B. (Abstract Scene - Opinion, 5.62%), I.D. (Abstract Scene - Talk bubbles, 5.19%), V.C. (Generic Scene - Nature Scene, 6.82%), V.D. (Generic Scene - Location with specific information, 6.92%), V.E. (Generic Scene - Location, 4.87%), VI.B. (Generic Object - Artificial object with location information, 5.49%), and VI.E. (Generic Object - Human with location information, 5.68%).

The distribution of the subcategories within the two most popular main categories of Q2 (index term) is consistent with that of Q1. The most popular subcategories among the two categories were: V.F. (Generic Scene - Event, 12.74%), and VI.D. (Generic Object - Artificial object, 13.88%). The next most frequently chosen subcategories are: V.E. (Generic Scene - Location, 8.48%), V.C. (Generic Scene - Nature scene, 8.25%), V.D. (Generic Scene - Location with specific information, 6.77%), VI.E. (Generic Object - Human with location information, 9.49%), VI.B. (Generic Object - Artificial object with location information, 9.46%), and VI.A. (Generic Object - Nature object, 5.53%). Many participants described "location" information in both Q1 and Q2. It seems natural that the participants assigned location information for V, Generic Scene; however, they also provided such information for VI, Generic Object. In contrast, the percentages of chosen subcategories related to color description are low. VI.C. (Generic Object - Artificial object with color description), was chosen 0.84% of the time for Q1, and 1.68% for Q2. VI.H. (Generic Object - Living thing with color description), was rated 1.75% for Q1, and 2.62% for Q2.

Some subcategories in the HOPR model were chosen very infrequently. Combining the results from Q1 and Q2, Abstract Object, Specific Scene (Nature scene with location information), Physical content (Visual extension), Specific Object (Object with specific name), Physical content (Geometric primitives), and Specific Scene (Events with location information) were selected less than 5 times each.

### 5.1.1 Comparison with Pyramid and Nine classes models

Applying the results of this experiment to the Pyramid model, we may say that most chosen entry points fit into level 5 (Generic Objects) and level 6 (Generic Scene) in both describing and indexing the image. The results did not correspond much with level 2 (Global Distribution), 7 (Specific Objects), 8 (Specific Scene), or 9 (Abstract Objects). Using the HOPR model, level 10 (Abstract Scene) was ranked as the second most popularly chosen category in describing the image.

Relating Burford, Briggs, & Eakins' nine classes of image retrieval to the results presented here is not an easy task. It seems that Semantic units would be similar to HOPR's Generic Object. A combination of Contextual Abstraction, Cultural Abstraction, and Emotional Abstraction might best correspond with Abstract Scene. Semantic Units, Contextual Abstraction, and Cultural Abstraction would relate to Generic Scene. As mentioned above, Neal [19, 25] utilized an adaptation of this model, and found that photojournalism professionals preferred searching most by events happening in the photo, objects represented in the photo, UADs, certain metadata about the image itself (such as the date on which the picture was shot), and browsing. The top two choices (events and objects) are similar to those most frequently chosen in the current study, despite the different user groups and contexts.

**Hyuk-Jin Lee and Diane Neal**

*5.2. Popular basic levels and index terms per image type*

Next, the two or three most frequent categories for each of the eight image types we identified through Flickr tag classification are shown in Table 3. First, the data shows that there is not much difference between Q1 and Q2 in terms of the most frequent subcategories in all image types. Five types match both the first and the second most frequent subcategories: Color, Concept, Environment, Event, and Object. The other three types also have identical subcategories in which only their ranks are changed between the first and the second frequent: Living Thing, Places, and Time. The major difference between the two questions is the existence of Abstract Scene subcategories. In Q1, five image types (Event, Object, Living Thing, Places, and Time) are described with Abstract Scene subcategories such as “Talk bubbles”, “Emotion”, “Assumption”, or “Opinion,” unlike Q2.

*Table 3. Most frequently chosen subcategories per image type for Q1 and Q2*

	Color	Concept	Environment	Event	Living Thing	Object	Places	Time
<b>Q1</b>	V.C. 126 (29.3%)	V.F 101 (23.5%)	VI.D. 95 (22.1%)	V.F. 150 (34.9%)	VI.G. 83 (19.3%)	VI.D. 117 (27.2%)	V.D. 89 (20.7%)	V.A. 70 (16.3%)
	VI.A. 90 (20.9%)	V.D. 91 (21.2%)	V.C. 67 (15.6%)	I.D. 73 (17%)	VI.E. 66 (15.3%) I.A. 51 (11.9%)	VI.B. 83 (19.3%) I.A. 55 (12.8%)	V.E. 75 (17.4%) I.A. 58 (13.5%)	VI.E. 69 (16%) I.E. 35 (8.1%) I.B. 34 (7.9%)
<b>Q2</b>	V.C. 155 (35.2%)	V.F.143 (32.5%)	VI.D. 106 (24.1%)	V.F.16 2 (36.8%)	VI.E. 92 (20.9%)	VI.D. 175 (39.8%)	V.E. 143 (32.5%)	VI.E. 109 (24.8%)
	VI.A. 111 (25.2%)	V.D. 94 (21.4%)	V.C. 89 (20.2%)	* I.D. 27	VI.G. 88 (20.0%)	VI.B. 174 (39.5%)	V.D. 72 (16.4%)	* V.A. 38
			VI.E. 86 (19.5%)		I.A. 58 (13.5%)			

For Color photographs, the most frequent categories are Generic Scene (Nature scene) and Nature object (Ordinary nature object). Interestingly, there was no ranking for Physical Content (Color and light). For the Concept photos, the most frequent categories are Generic Scene (Event) and Generic Scene (Location with specific information). Abstract Scene or Abstract Object, with which one might think has a close relationship with Concept, is not frequently chosen. For the Environment pictures, Generic Object (Artificial object) and Generic Scene (Nature scene) are frequently chosen. Again, Generic Scene (Environment) is not ranked. For the Event images, the two categories most frequently chosen are Generic Scene (Event) and Abstract Scene (Talk bubbles). For this category, the participants chose the Event category (in Generic Scene); however, there is no ranking for any other Event-related category such as Specific Scene (Event with location information). For the Living Thing images, the chosen categories are mostly related to the image type: Generic Object (Living thing with location information), Generic Object (Human with location information), and Abstract Scene (Emotion). The participants focused on “location information” for Human or Living Thing more than any other type of information. Next, for the Object type, the chosen categories are again related to the image type. The coders chose Generic Object (Artificial object) and Generic Object (Artificial object with location information). It was the same for Place pictures; Generic Scene (Location with specific information) and Generic Scene (Location) are the most frequently chosen categories. Finally, for Time photographs, Generic Object (Nature object- ordinary nature

## Hyuk-Jin Lee and Diane Neal

object), Generic Object (Artificial object), and Generic Object (Artificial object with color description) are the most frequent categories.

### 5.3. Popular basic levels and index terms per image

Comparisons between our eight “Flickr tag” categories and the most frequently chosen categories create several interesting results (see Table 4). First, similar results were found between Q1 and Q2. Among the 16 images, 11 images have exact matches between Q1 and Q2. One image demonstrates a simple rank change between the most frequent and the second most frequent. Only 4 images have conflicting results: Color (Sunset), Event (Birthday), Places (Pool), and Time (Street with snow). While the participants described the image of Color (Sunset) in Q1 as Generic Scene (Nature scene) based on the coding, they chose Generic Object (Artificial object with color description) when they indexed it in Q2. When the subjects described the image of Concept (Birthday), they stated what the people in the image might be thinking or saying (Abstract Scene - talk bubbles). However, when the participants indexed the same image in Q2, they most frequently chose Generic Object (Artificial object). In the Places-Pool image, the participants described it as Generic Object (Artificial object with location information) and Abstract Scene (Emotion) in Q1. However, they do not agree strongly with the index terms (Q2), and there is no particular subcategory agreement. Finally, while the participants described the image of Time (Street with snow) as Generic Object (Nature object) the most, they indexed the image as Generic Object (Artificial object).

Table 4. Most frequently chosen subcategories per image for Q1 and Q2

	Color (Sunset)	Concept (Cheers)	Environment (Cloud)	Event (X-mas deco.)	Living Thing (Puppy)	Object (Water fountain)	Places (City view)	Time (Shore)
<b>Q1</b>	V.C. 123 (53.9%)	V.F. 75 (32.9%) I.A. 35 (15.4%) I.E. 32 (14%)	V.C. 67 (29.4%)	V.F. 98 (43%) I.A. 22 (9.6%) I.B. 21 (9.2%)	VI.G. 83 (36.4%) I.A. 50 (21.9%) VI.I. 40 (17.5%)	VI.B. 80 (35.1%) I.A. 44 (19.3%)	V.D. 82 (36%) V.E. 65 (28.5%)	VI.E. 69 (30.3%) I.E. 35 (15.4%)
<b>Q2</b>	VI.C. 155 (67.4%)	V.F. 89 (38.7%) VI.E. 36 (15.7%)	V.C. 83 (36.1%) VI.A. 48 (20.9%)	V.F. 116 (50.4%)	VI.G. 88 (38.3%) VI.I. 77 (33.5%)	VI.B. 118 (51.3%) VI.D. 65 (28.3%)	V.E. 127 (55.2%) V.D. 60 (26.1%)	VI.E. 109 (47.4%) V.E. 49 (21.3%)
	Color (Green leaves)	Concept (Construction field)	Environment (Sun over a statue)	Event (Birthday)	Living Thing (Children)	Object (Platform)	Places (Pool)	Time (Street with snow)

Hyuk-Jin Lee and Diane Neal

Q1	VI.A. 90 (39.5%)	V.D. 91 (39.9%)	VI.D. 95 (41.7%)	I.D. 63 (27.6%)  V.F. 51 (22.8%)	VI.E. 66 (28.9%)	VI.D. 81 (35.5%)  I.D. 25 (11%)	VI.B. 63 (27.6%)  I.A. 49 (21.5%)	VI.A. 62 (27.2%)
Q2	VI.A.11 1 (48.3%)  VI.H. 67 (29.1%)	V.D. 90.4 (40.9%)  V.F. 54 (23.5%)	VI.D.106 (46.1%)  VI.B. 86 (37.4%)	VI.D. 57 (24.8%)  VI.F. 48 (20.9%)  V.F. 46 (20.0%)	VI.E. 92 (40.0%)	VI.D. 110 (47.8%)  VI.B. 56 (24.3%)	Nothing peculiar V.E. 16 (7%)	VI.D. 45 (19.6%)  VI.C. 39 (17%)  V.A. 38 (16.5%)

There is not a single agreement regarding the most frequent subcategory between two images under the same image type, such as Color or Time, in the results for either question (see Table 4). The results may imply many things. First, although we concede that it was a subjective process for the authors to categorize the images, the results indicate how difficult is for a few experts or indexers to categorize the images with agreement. In other words, this demonstrates how different chosen entry points or index terms may be between different users for the same image. This outcome may support the necessity of application of Web 2.0 concept in recognizing the entry points or indexing the online images, which allows using the collective knowledge on such a subjective area.

Second, the General Object category should be further investigated. Only two categories' images in the seven main categories match with the given categories: Object and Living Thing (both are under General object, main category). But in many other categories, when there are two conflicting results between the two images under the same type, the most frequent category found is Generic Object. For example, in two Event images, the Christmas decoration picture is most often categorized as Generic scene (Event) in both Q1 and Q2, but the chosen subcategories for the Birthday image are Abstract Scene (Talk bubbles) in Q1, and Generic Object (Artificial object) in Q2. In two Environment images, the Cloud image is categorized most often as Generic Scene (Nature scene) in both questions, and the "Sunshine over a Statue" photo is categorized most often as Generic Object (Artificial object or Human) in both questions. The results indicate that the size or location of the object (either Artificial object or Human) are not the main factors in end user's image decision, but the context around the object is a more influential factor. For example, even though Human was not chosen as the most important category in Concept (Cheers) or Event (Christmas decoration), the actual size of the human in such image types is bigger than that in Living Thing (Children) or Time (Shore) types where Human was recognized as the important category. Also, some images show that artificial objects are recognized often regardless of their sizes, and the image is categorized accordingly. For example, Event (Birthday) and Environment (Sunshine over a statue) are indexed as Generic Object (Artificial object). Thus, our results show the importance of further investigation on what factors influence the end user's decision in regards to whether the image is about a Generic Object, such as an Artificial object or a Human. For example, the facial expression or body posture of a human, living thing or artificial object may be the powerful factor for the viewer's decision [14].

## Hyuk-Jin Lee and Diane Neal

### 5.3.1 Comparison with Pyramid and Nine classes models

In relation to the Pyramid model, the results show that different image types have different popularly chosen subcategories that belong to the same main categories. For example, the most frequently chosen subcategory for Color in Q1 is Generic Scene (Nature scene), and the most frequent one for Concept is Generic scene (Event). Even though both subcategories are under the same category (Generic Scene), they are to be treated differently for more effective classification, since they are distinct categories. The definition of Generic Scene in the Pyramid model, or 'scenes at their most general level of description' [37], is perhaps too broad to be useful in the representation, search, and retrieval of Web images. Thus, considering the distribution of different subcategories according to the image types, it seems useful to subdivide the Pyramid model into more refined subcategories in order to reflect the actual entry points of Web images better, especially for the most frequently used categories. Even though the HOPR model is not complete yet, such an attempt is providing useful subcategories.

Using these results, it turned out to be difficult to compare results with or classify the images based on the Pyramid model or the nine classes of image content, because the participants tended to choose entry points and UADs belonging to subcategories that reside in the HOPR main categories. Based on the results of this study, the models under discussion are perhaps not constructed to represent image content at users' basic levels of description.

### 5.4. Utility of the HOPR model

Regarding the utility of the HOPR model, it is too soon to make definitive conclusions. For example, even though our coding was a viable approach following qualitative thematic coding principles, certain HOPR model categories drew very few to no indexing applications by the test subjects. We may have to consider removing those categories and therefore, this model should be refined based on further research. However, providing subcategories under the top three main categories utilized by participants in this study would provide an applicable range of common choices for assigning guided UADs. While most responses fell in HOPR's mid-range in terms of description levels, the third most popular level was Abstract Scene, the most abstract level of the model. Again, it is important to understand the wide spectrum of possible description levels provided online for personal photographs. However, based on the data in this study, the basic level of description falls mainly within the categories of Generic Scene and Generic Object. Since these categories fall within the middle range of abstraction within HOPR, it could be said that our model adequately represents various subordinate, superordinate, and basic level. In the model, there are more specific subordinate categories as well as more abstract superordinate categories surrounding the basic level of Generic Scene and Generic Object.

## 6. Conclusions and future research

Although the number of images and the participants in this study may not be numerous enough to reach any definitive conclusions, and the subcategories need to be refined, the results are provocative in terms of discussion and future research. Our findings show that for general Web-based photographs, 1) popular image description models do not adequately represent the various levels of likely UADs, 2) there are popular main categories and subcategories that need to be investigated and further developed, and 3) the development of a model that can satisfactorily assist in the representation of Web-based photographs from the users' point of view must continue. We do not argue that the Pyramid or the nine classes of image content are not useful; instead, we suggest focusing on the most useful levels and extending or refining them in order to make the models more useful for the general set of online photograph users. Also, due to the differences in results between the entry points and indexing terms, we find it necessary to explore ways to bridge the gap between users' basic levels or entry points for images and UADs as well as traditional indexing terms. This need is seemingly not as complex in the text retrieval domain as it might be in non-text retrieval.

## Hyuk-Jin Lee and Diane Neal

Next, the results demonstrate that while the subjects actively described the category I (Abstract Scene) including emotion, opinion, talk bubbles, or assumptions about the images when they assigned entry points (Q1), they did not do so when they had to assign the index terms (Q2). This result could imply that Web image users first experience feelings that belong to the Abstract Scene category; however, we think they might assume that such feelings cannot be used as index terms for searching. Thus, they translate them as either Generic Object or Generic Scene for indexing purposes. It is an interesting result, and should be investigated more. A few related studies showed the difficulty in indexing the emotional aspects and the necessity of emotional index category. Rorissa & Hastings [40] said that although it is theoretically possible to create emotional concepts in knowledge organization systems, it has not been actively applied in information practice due to a lack of experiences with emotional categories. Laine-Hernandez and Westman [41] also argued, "The influence of the photograph's emotional tone on categorization has not been discussed much in previous studies." The Abstract Scene category seems important in image indexing based on this study, and it is necessary to understand the reasons that hinder people to translate their feelings into index terms. The area of Emotional Information Retrieval (EmIR) is just now starting to be explored [42-44].

We found that among the 10 levels in the Pyramid model, level 5 (Generic Objects) and level 6 (Generic Scene) are mostly used in both describing the image entry points and in indexing Web-based photographs. These levels possibly represent the "basic levels" of abstraction within our model. A major difference between the results of the two questions was level 10 (Abstract Scene), which was the second most frequently chosen level in describing the entry points. Most responses corresponded to Semantic units, Contextual abstraction, Cultural abstraction, and Emotional abstraction in the nine classes of image content. Therefore, further investigation should focus on these types of description. For example, it would be meaningful to study whether either image basic levels or UADs terms are more effective in image retrieval. Various distributions of HOPR's subcategories indicate that the Pyramid levels, especially Generic Object, Generic Scene and Abstract Scene, should be sub-classified to meet user needs. The subcategories and findings in this study may be applicable for such an agenda. Similar development is suggested for the nine classes of image content.

As introduced in the literature review, many studies have indicated the difficulty of indexing image Aboutness due to its subjective characteristics. The findings in this study emphasize the importance of a "down-top" approach instead of a "top-down" approach; it is impossible for a few experts/indexers to adequately classify the scores of images online due to time and subjective Aboutness. Fortunately, the currently popular act of tagging photographs and other documents online is a "down-top" approach, allowing many users to contribute their knowledge [45].

While concerns related to a lack of authority control exist within the library and information science research and practice communities, the popularity of participating in tagging photographs online is undeniable. It seems unreasonable to expect users to begin using traditional methods of description after experiencing the UAD approach. However, perhaps the HOPR model, and its continued development, can be a step toward understanding user behaviour when searching for and describing photographs online. This understanding will enable professionals to develop or enhance systems that allow UADs while providing users some contextualised direction toward a balanced use of authority control [25]. As a related future study, it would be intriguing to compare the results by applying the Web image categories developed by image experts with the findings of this study. Also, we plan to validate the results with more images and various types of users.

As a related further analysis, we may compare the statistical results between Q1 and Q2 in terms of the most frequent subcategories per image, type, and total data in order to investigate any particular pattern. In addition, future studies might investigate how to revise the Levels 5, 6, and 10 of the Pyramid Model, Levels I, V, and VII of the HOPR model, and corresponding classes of image content [5] with subcategories in order to be practically useful in Web-based image information retrieval. For example, if image users can assign entry points or index terms using tagging combined with useful hierarchical structures, online image access may improve. A hierarchy such as the HOPR model, and an understanding and utilization of the basic levels present in the hierarchy, would allow for some authoritative guidance as well as for browsing. The hierarchy could reflect users' knowledge of the indexed photographs over time as they add to the system [46]. The architecture could involve methods beyond the simple hierarchy; for example, perhaps facets, visualizations, or Semantic Web approaches could be incorporated. Many opportunities for research and development exist in this area.

## Hyuk-Jin Lee and Diane Neal

### References

- [1] P. Enser, The evolution of visual information retrieval, *Journal of Information Science* 34(4) (2008) 531-546.
- [2] E. Panofsky, *Meaning in the visual arts* (NY: Doubleday Anchor Books, Garden City, 1955).
- [3] S. Shatford, Analyzing the subject of a picture: A theoretical approach, *Cataloging & Classification Quarterly* 6(3) (1986) 39-62.
- [4] P. Enser, Visual image retrieval: seeking the alliance of concept-based and content-based paradigms, *Journal of Information Science* 26(4) (2000) 199-210.
- [5] B. Burford, P. Briggs, and J.P. Eakins, A taxonomy of the image: On the classification of content for image retrieval, *Visual Communication* 2(2) (2003) 123-161.
- [6] M.G. Krause, Intellectual problems of indexing picture collections, *Audiovisual Librarian* 14 no. 4 (November) (1988) 73-81.
- [7] H. Greisdorf and B. O'Connor, Modelling what users see when they look at images: A cognitive viewpoint, *Journal of Documentation* 58(1) (2002) 6-29.
- [8] A. Goodrum and A. Spink, Image searching on the Excite Web search engine, *Information Processing and Management* 37(2) (2001) 295-311.
- [9] B.J. Jansen, Searching for digital images on the web, *Journal of Documentation* 64(1) (2008) 81-101.
- [10] M. Hogan., C. Jörgensen and P. Jörgensen, The Visual Thesaurus in a Hypermedia Environment: A Preliminary Exploration of Conceptual Issues and Applications. In: International Conference on Hypermedia & Interactivity in Museums (ICHIM 1991): *Proceedings of International Conference on Hypermedia and Interactivity in Museums* (pp. 202-221), (1991).
- [11] P. Enser, Progress in documentation: Pictorial information retrieval, *Journal of Documentation* 51(2) (1995) 126-170.
- [12] B.C. O'Connor and R.B. Wyatt, *Photo provocations* (MD: The Scarecrow Press, Lanham, 2004).
- [13] H. Besser, Visual access to visual images: The UC Berkeley Image Database Project, *Library Trends* 38(4) (1990) 787-798.
- [14] H.F. Greisdorf and B.C. O'Connor, *Structures of image collections: From Chauvet-Pont-D'Arc to Flickr* (CT: Libraries Unlimited, Westport, 2008).
- [15] M.E. Rorvig, C.H. Turner, and J. Moncada, The NASA Image Collection Visual Thesaurus, *Journal of the American Society for Information Science* 50(9) (1999) 794- 798.
- [16] P. Wilson, Searching: Strategies and evaluation. In H. D. White, M. J. Bates, & P. Wilson (Eds.), *For information specialists* (pp. 153-181) (NJ: Ablex Publishing Corporation, Norwood, 1992).
- [17] L. Standing, J. Conezio, and R.N. Haber, Perception and memory for pictures: Single-trial learning of 2500 visual stimuli, *Psychonomic Science* 19(2) (1970) 73-74.
- [18] S. Ornager, Image retrieval: theoretical analysis and empirical user studies on accessing information in images. In C. Schwartz & M. Rorvig (Eds.), *Proceedings of the 60th ASIS Annual Meeting* (pp. 202-211) (NJ: Information Today, Medford, 1997).
- [19] D.R. Neal, *News photography image retrieval practices: Locus of control in two contexts* (Unpublished manuscript, University of North Texas, Denton, TX, 2006).
- [20] K.K. Matusiak, Information seeking behavior in digital image collections: A cognitive approach, *The Journal of Academic Librarianship* 32(5) (2006) 479-488.
- [21] K. Markey, Interindexer consistency tests: a literature review and report of a test of consistency in indexing visual materials, *Library and Information Science Research* 6 (1984) 155-177.

**Hyuk-Jin Lee and Diane Neal**

- [22] N. Conduit and P. Rafferty, Constructing an image indexing template for The Children's Society, *Journal of Documentation* 63(6) (2007) 898-919.
- [23] M.J. Bates, Indexing and access for digital libraries and the Internet: Human, database, and domain factors, *Journal of the American Society for Information Science and Technology* 49(13) (1998) 1185-1205.
- [24] J. Trant, Exploring the potential for social tagging and folksonomy in art museums: Proof of concept, *New Review of Hypermedia and Multimedia* 12(1) (2006) 83-105.
- [25] D. Neal, News photographers, librarians, tags, and controlled vocabularies: Balancing the forces, *Journal of Library Metadata* 8(3) (2008) 199-219.
- [26] J.W. Yoon, Towards a user-oriented thesaurus for non-domain-specific image collections, *Information Processing and Management* 45 (2009) 452-468.
- [27] S. Yang, S.K. Kim, K.S. Seo, Y.M. Ro, J-Y. Kim, and Y.S. Seo, Semantic categorization of digital home photo using photographic region templates, *Information Processing and Management* 43 (2007) 503-514.
- [28] P.G.B. Enser, C.J. Sandom, J.S. Hare, and P.H. Lewis, Facing the reality of semantic image retrieval, *Journal of Documentation* 63(4) (2007) 465-481.
- [29] C-F. Tsai, A review of image retrieval methods for digital cultural heritage resources, *Online Information Review* 31(2) (2007) 185-198.
- [30] P. Jolicoeur, M.A. Gluck, and S.M. Kosslyn, Pictures and names: Making the connection, *Cognitive Psychology* 16 (1984) 243-275.
- [31] E. Rosch, C.B. Mervis, W.D. Gray, D.M. Johnson, and P. Boyes-Braem, Basic objects in natural categories, *Cognitive Psychology* 8 (1976) 382-439.
- [32] A. Wilhelm, Y. Takhteyev, R. Sarvas, N. Van House, and M. Davis, Photo annotation on a camera phone. In: Extended Abstracts of the 2004 Conference on Human Factors in Computing Systems (CHI 2004): *Proceedings of the Conference on Human Factors in Computing Systems* (pp. 1403 – 1406) (ACM Press, New York, NY, 2004).
- [33] A. Rorissa, User-generated descriptions of individual images versus labels of groups of images: A comparison using basic level theory, *Information Processing and Management* 44 (2008) 1741-1753.
- [34] A. Rorissa and H. Iyer, Theories of cognition and image categorization: What category labels reveal about basic level theory, *Journal of the American Society for Information Science and Technology* 59(9) (2008) 1383-1392.
- [35] B. Stvilia and C. Jörgensen, User-generated collection-level metadata in an online photo-sharing system, *Library & Information Science Research* 31(1) (2009) 54-65.
- [36] A. Jaimes and S.-F. Chang. A conceptual framework for indexing visual information at multiple levels. In: G. Beretta & R. Schettini (eds), *IS&T/SPIE Conference on Internet Imaging 2000: Proceedings of the First IS&T/SPIE Internet Imaging 2000* (pp. 3964), (2000).
- [37] C. Jörgensen, A. Jaimes, A.B. Benitez, and S.-F. Chang, A conceptual framework and empirical research for classifying visual descriptors, *Journal of the American Society for Information Science and Technology* 52(11) (2001) 938-947.
- [38] P. Enser, Visual Image Retrieval, *Annual Review of Information Science and Technology* 24(1) (2008) 1-42.
- [39] J.R. Landis and G.G. Koch, The measurement of observer agreement for categorical data, *Biometrics* 33(1977) 159-174.
- [40] Rorissa, A., & Hastings, S.K., Free sorting of images: Attributes used for categorization. In *Proceedings of the 67th Annual Meeting of the American Society for Information Science and Technology (Vol. 42). Joining Research and Practice: Social Computing and Information Science* (pp. 360–366), (Medford, NJ, 2004).









---

**Hyuk-Jin Lee and Diane Neal**









- [41] Laine-Hernandez, M., & Westman, S., Image semantics in the description and categorization of journalistic photographs. In *Proceedings of the 69th Annual Meeting of the American Society for Information and Technology (Vol. 44). Joining Research and Practice: Social Computing and Information Science*, (Medford, NJ, 2006)
- [42] H.J. Lee and D. Neal, Towards web 2.0 music information retrieval: Utilizing emotion-based, user-assigned descriptors. In *Proceedings of the 70th Annual Meeting of the Annual Meeting of the American Society for Information Science and Technology (Vol. 45). Joining Research and Practice: Social Computing and Information Science* (pp. 732–741), (Medford, NJ, 2007).
- [43] S. Schmidt and W. Stock, Collective indexing of emotions in images: A study in emotional information retrieval, *Journal of the American Society for Information Science and Technology* 60(5) (2009) 863-876.
- [44] D. Neal, A. Campbell, J. Neal, C. Little, A. Stroud-Mathews, S. Hill, and C. Bouknight-Lyons, Musical facets, tags, and emotion: Can we agree? In: 2009 iConference: *Proceedings of the 2009 iConference – iSociety: Research, education, engagement*, (2009).
- [45] J. Surowiecki, *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nation* (New York: Doubleday, 2004).
- [46] B.C. O'Connor, *Explorations in indexing and abstracting: pointing, virtue, and power* (CO: Libraries Unlimited, Englewood, 1996).

Hyuk-Jin Lee and Diane Neal

Appendix I: Personal photographs used in the study

Color 1 (Sunset)	Environment 1 (Cloud)
	
Living thing 1 (Puppy)	Objects 1 (Water fountain)
	
Places 1 (City view)	Events 1 (X-mas decoration)
	

Hyuk-Jin Lee and Diane Neal

<p>Concepts 1 (cheers)</p>	<p>Time 1 (Shore)</p>
	
<p>Color 2 (Leaves)</p>	<p>Concepts 2 (Construction field)</p>
	
<p>Places 2 (Pool)</p>	<p>Events 2 (Birthday)</p>
	
<p>Living thing 2 (Children)</p>	<p>Environment 2 (Sun over a statue)</p>
	

---

**Hyuk-Jin Lee and Diane Neal**

Objects 2 (Platform)	Time 2 (Street with snow)
	

## Hyuk-Jin Lee and Diane Neal

### Appendix II: Hierarchy for Online Photograph Representation (HOPR)

I. **Abstract scene:** what the image as a whole represents, and may be very subjective (e.g., sadness, happiness, power, heaven, and paradise)

A. Emotion – the image makes the user feel emotion, or the picture conveys an emotion to the user

*"Another big city skyline – where nature is virtually blocked out – not a happy prospect but one that is common"*

*"A happy party! kind of messy"*

B. Opinion – the user states an opinion about what he or she sees in the image

*"It's the city. I don't like big cities."*

*"It is nice to see some trees in the city"*

C. Transference – the picture makes the user think about something in his or her own life, even if it is not related

*"Cheers! Jim and Betty at our last party!"*

*"drinking and partying at the Church fellowship!"*

D. Talk bubbles – the user states what he or she thinks the subject(s) in the image might be thinking or saying

*"Brrr! This water's freezing!"*

*"Now, which one should I use?"*

E. Assumption – the user assumes something is true about the image, regardless of whether he or she knows the actual facts

*"A man on vacation at the beach, probably not from close by"*

*"He doesn't want to get in the water because it's cold"*

II. **Abstract object:** specialized or interpretative knowledge about what the objects represent. It is completely subjective, and assessments between different users may vary greatly (e.g., a woman in a picture shows anger)

III. **Specific scene:** attributes that carry specific knowledge about the scene (e.g., Paris, Time Square, Central Park, etc.)

A. Events with location information - it is obvious that the event information is correct. A user is specifically focusing on the location of that event.

*"People drinking at a church party"*

*"An office holiday party at a friend's house"*

B. Specific location - the user knows about the picture location.

*"Fort Worth, Texas, skyline downtown view from west"*

*"Fort Worth Skyline"*

C. Nature scene with location information- the user knows about the location where the nature scene was captured.

*"Sunrise over Clarewood Park"*

*"Summer thunderstorm on the day before our vacation in Fort Worth"*

IV. **Specific object:** identified and named objects. Specific knowledge of the objects in the image is required. (e.g., Bill Clinton, Eiffel Tower).

## Hyuk-Jin Lee and Diane Neal

### A. Object with specific name- specific object name

*"Betty Utter memorial birdbath in Botanical Gardens, Fort Worth"*

*"pier 1 building"*

## V. **Generic scene:** scenes at their most general level of description (e.g., city, landscape, indoor, outdoor, still life, and portrait)

### A. Environment (e.g., Darkness, night cold, snowy)

### B. Nature scene with specific description- Nature scene with a specific description such as geographic information or weather condition

*"a beach scene on a clear, somewhat windy day indicated by the whitecaps rolling toward shore...one man walking toward the beach, two other people wading out into the water which is at thigh depth"*

*"looking over a line of trees at the skyline of a large city"*

### C. Nature scene- everyday life nature scene

*"Sunset through trees"*

*"cloud formation"*

*"tree silhouette cumulous cloud"*

### D. Location with specific information- location information with specific information such as time or geographic information

*"City skyline from a nearby hill"*

*"A wintry downtown skyline all blue and grey with trees in the foreground"*

### E. Location- location information

*"In the park"*

*"city scene, buildings skyline"*

### F. Event- it is obvious that the event information is correct.

*"Christmas tree decoration"*

*"ocean activities"*

## VI. **Generic object:** the most general level of object description, which can be recognized with everyday knowledge (e.g., apple, man, chair).

### A. Nature object- ordinary nature object

*"clouds"* (this example could also be in Nature scene though)

### B. Artificial object with location information – artificial object with a specific location description

*"fountain in a state park"*

*"Photo of fountain in park along walking trail"*

### C. Artificial object with color description

*"A dark brown leather sofa with a little pup napping on a black and brown pillow"*

### D. Artificial object- ordinary artificial object

*"birdbath or fountain"*

*"woodland setting"*

### E. Human with location information- human with a specific location description

---

## Hyuk-Jin Lee and Diane Neal

*“Tourist on a beach”*

*“photographers in the park”*

F. Human – human being or human vocation

*“Tourist”*

*“beach vacationer”*

G. Living thing with location information- Living thing with a specific location description

*“small dog sleeping on sofa pillow”*

*“small dog in brown chair”*

H. Living thing with color description – Living thing with its color description

*“a small red-brown dog curled up on a small patterned pillow on a brown leather couch or chair”*

I. Living thing – ordinary living thing

*“toy breed dog sleeping”*

*“sleeping puppy”*

### VII. Physical content

A. Visual extension (depth, shadow, perspective)

B. Geometric primitives (lines and shapes)

C. Perceptual primitives (color and light)